



# DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

## An objective function exploiting suboptimal solutions in metabolic networks

The Harvard community has made this article openly available.  
[Please share](#) how this access benefits you. Your story matters.

<b>Citation</b>	Wintermute, Edwin H, Tami D Lieberman, and Pamela A Silver. 2013. "An objective function exploiting suboptimal solutions in metabolic networks." <i>BMC Systems Biology</i> 7 (1): 98. doi:10.1186/1752-0509-7-98. <a href="http://dx.doi.org/10.1186/1752-0509-7-98">http://dx.doi.org/10.1186/1752-0509-7-98</a> .
<b>Published Version</b>	<a href="https://doi.org/10.1186/1752-0509-7-98">doi:10.1186/1752-0509-7-98</a>
<b>Accessed</b>	February 16, 2015 11:53:34 AM EST
<b>Citable Link</b>	<a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:12406822">http://nrs.harvard.edu/urn-3:HUL.InstRepos:12406822</a>
<b>Terms of Use</b>	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a>

*(Article begins on next page)*

RESEARCH ARTICLE

Open Access

# An objective function exploiting suboptimal solutions in metabolic networks

Edwin H Wintermute<sup>1\*</sup>, Tami D Lieberman<sup>1</sup> and Pamela A Silver<sup>1,2\*</sup>

## Abstract

**Background:** Flux Balance Analysis is a theoretically elegant, computationally efficient, genome-scale approach to predicting biochemical reaction fluxes. Yet FBA models exhibit persistent mathematical degeneracy that generally limits their predictive power.

**Results:** We propose a novel objective function for cellular metabolism that accounts for and exploits degeneracy in the metabolic network to improve flux predictions. In our model, regulation drives metabolism toward a region of flux space that allows nearly optimal growth. Metabolic mutants deviate minimally from this region, a function represented mathematically as a convex cone. Near-optimal flux configurations within this region are considered equally plausible and not subject to further optimizing regulation. Consistent with relaxed regulation near optimality, we find that the size of the near-optimal region predicts flux variability under experimental perturbation.

**Conclusion:** Accounting for suboptimal solutions can improve the predictive power of metabolic FBA models. Because fluctuations of enzyme and metabolite levels are inevitable, tolerance for suboptimality may support a functionally robust metabolic network.

**Keywords:** Metabolism, Variability, Metabolic flux analysis, Networks

## Background

### Predicting metabolism with constraint-based models

Natural selection, acting within the laws of physics and chemistry, produces systems well adapted to local conditions. This basic evolutionary insight is formalized in constraint-based modeling. A biological system might be understood as the solution to a precisely formulated mathematical problem of optimization under constraint [1,2].

A constraint-based approach is often used to model metabolism in *E. coli*, for which nearly complete knowledge of the metabolic reaction network is available [3,4]. The exact stoichiometry of each biochemical reaction imposes a conservation-of-mass constraint that must hold in the steady state. In some cases thermodynamic, regulatory, or other constraints may be added to further restrict the possible metabolic reaction fluxes [5,6]. With the allowed flux space defined, common approaches select a flux vector that is optimal according

to some hypothesized objective function [7]. This optimal solution represents an *a priori* prediction of all cellular metabolic fluxes.

The objective most often chosen for microbial models is the maximization of growth rate or yield, an approach we call simply Flux Balance Analysis (FBA). The utility of FBA relies on the assumption that growth rate approximates overall fitness and is the primary focus of selection. Numerous successful applications attest to the value and versatility of FBA for predicting growth rates in a variety of contexts [8-11]. However, standard FBA formulations face practical and principal limitations.

In practice, FBA generally cannot predict a unique rate for all fluxes. A solution which maximizes growth rate is typically mathematically degenerate, describing a region in flux space rather than a single point. Solution degeneracy is a well-described problem in systems in systems like metabolism which are flexible, internally redundant, and underdetermined by data [12]. Applications of FBA are therefore complicated when predictions are required for fluxes other than growth.

In principal, metabolism cannot function only to maximize growth rate. This is evidenced by deletions of

\* Correspondence: ehwintermute@gmail.com; Pamela\_Silver@hms.harvard.edu

<sup>1</sup>Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

<sup>2</sup>Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA

Full list of author information is available at the end of the article

metabolic genes in *B. subtilis*, some of which cause increased growth rates and biomass yields relative to wild type [13]. Selective pressure to increase growth rate must be balanced by other demands on metabolism. Investments in cellular maintenance, sensory apparatus, osmoregulation, intercellular communication, or motility may reduce growth rate while improving overall fitness. Finally, limitations of evolutionary time and genetic variability may mean that metabolism is simply imperfect, and not optimal for any objective [2]. Thus we can not necessarily exclude from consideration the many flux configurations that support, for example, 90% maximal growth.

We address both difficulties with a revision of the metabolic objective at the heart of genome-scale metabolic optimization models. We propose that microbial metabolism is better represented as a cloud of nearly optimal flux distributions, rather than a single ideal and fixed solution. Under this hypothesis, regulation drives cellular fluxes to within a degenerate optimal region. Solutions within this region are not further optimized and considered equally plausible. We present a mathematical formulation of this theory that preserves the properties of continuous and convex optimization that make standard FBA mathematically elegant and computationally quick.

Our model allows stronger predictions in some cases, despite strictly weakening the optimality assumption that grounds standard FBA. The Perturbed Solution Expected Under Degenerate Optimality (PSEUDO) outperforms comparable methods in predicting the redistribution of central carbon fluxes that occurs in metabolic mutants of *E. coli*. Our model attributes central metabolism with significant flexibility to negotiate the trade-off between optimizing growth rate and matching a target flux vector. This is particularly relevant for metabolic engineering applications when genetically modified cells must adapt to suboptimal flux profiles.

The success of the PSEUDO method suggests that degenerate optimality may be a basic organizing principle for metabolism. In support of this hypothesis, we will show that reported measures of flux variability correlate with the dimensions of the degenerate optimal region. More degenerate fluxes, which theoretically do not require precise values for fast growth, indeed exhibit more variation. Tolerance for numerous optimal and near-optimal flux configurations would be an asset to microbial metabolism, enabling robust growth in the face of perturbations to the network. A degenerate organization of metabolism may be an essential adaptation given fundamental physical constraints on the ability of cells to control their internal and external environments [14].

## Results and discussion

### Common approaches to genome-scale optimization: FBA and MOMA

Figure 1 presents a geometric interpretation of the commonly used FBA and MOMA objective functions, contrasting them with the PSEUDO objective that we will describe below.

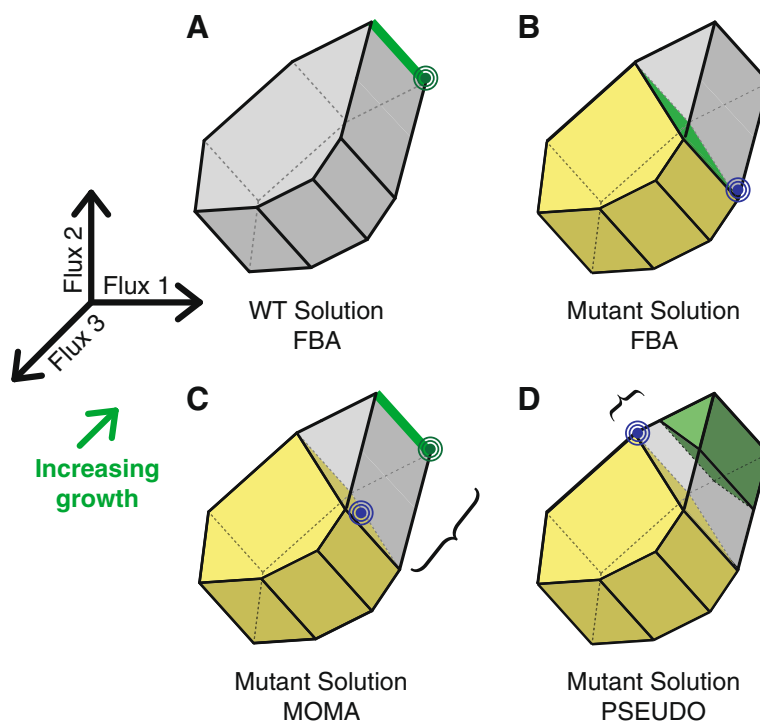
Standard FBA solves for a vector of metabolic fluxes within constrained flux space that maximizes cellular growth rate (Figure 1A). Because time is not represented in the model, predictions of growth rate or yield are treated identically. Mutations within the FBA framework are modeled as additional constraints that remove a region of the allowed flux space. For example, deletion of the *pgi* gene in *E. coli* eliminates phosphoglucose isomerase activity and constrains flux through that reaction to be zero. A mutant model is then re-solved to predict a new growth rate optimum (Figure 1B).

A popular alternative method for predicting mutant behavior is based on the Minimization of Metabolic Adjustment (MOMA) [15]. A mutant may not grow optimally if natural selection has not had a chance to act on the new genetic background. Instead, MOMA hypothesizes that a mutant will tend to approximate the wild-type state as closely as possible. Formally, a MOMA flux vector is found with minimum Euclidean distance to a single optimal wild-type profile, subject to the constraints of mutation (Figure 1C). This method therefore requires as input a unique optimal wild-type flux vector, which may be known from empirical measurements. In practice however, this point is often predicted with a standard FBA model.

Both FBA and MOMA use convex objective functions and convex constraints. Applications of these models can therefore access a powerful suite of convex programming algorithms [16]. A flux vector identified by convex programming is guaranteed to be unique, globally optimal, and can be computed in milliseconds. This is unlike most nonlinear optimization methods, which are computationally intensive and often can not guarantee a global optimum. The fast run times provided by convexity means that thousands of model variants can be rapidly re-solved in seconds on a conventional desktop. In metabolic engineering, for example, mutations can be screened combinatorially *in silico* for sets that improve production of a metabolite of interest.

### The problem of degeneracy in genome-scale models

Although solutions to FBA and MOMA problems are guaranteed to be globally optimum, they are not guaranteed to be unique. In practice, many different flux profiles allow equally optimum growth (Additional file 1: Figure S1 and Additional file 2: Figure S2). The problem



**Figure 1 FBA, MOMA and PSEUDO approaches to predicting metabolic fluxes.** (A) In FBA wild-type flux space is constrained to a polytope defined by thermodynamic and conservation-of-mass requirements. A linear objective describing cell growth, the green arrow, is maximized within this region. If the growth vector is perpendicular to a facet of the constrained polytope then a range of fluxes allow equally optimum growth, indicated by the heavy green edge. However, a linear programming solver can return only a single optimal point, the green target. (B) Mutations are represented as additional linear constraints that reduce the size of the allowed flux polytope. The yellow region represents the subset of wild-type fluxes allowed under a mutation. FBA finds a new optimum within this space as for the wild type. The green face represents a range of equally optimal mutant solutions. The blue target is a single point that a solver might return. (C) MOMA is an alternative approach for predicting mutant fluxes. The point in the mutant region, blue target, is found that minimizes the distance to a wild-type solution, green target. If FBA was used to generate the wild-type solution, then alternative optima may exist along the heavy green edge. (D) The PSEUDO strategy does not use FBA to select a wild-type flux vector. Instead we define a degenerate optimal region that contains all flux distributions capable of supporting near-maximal growth. A solution within the mutant region is found with minimum distance to this degenerate optimal region. Note that PSEUDO may select a point in mutant flux space different from the MOMA solution and closer to the growth-optimal region.

of degeneracy is encountered frequently in the literature, and numerous attempts have been made to address it.

Degeneracy can be reduced by further constraining the model using known regulatory interactions [17], metabolite concentrations [18] or thermodynamic laws [19]. Previously measured flux rates, when available, are a particularly valuable guide for further predictions [15,20,21].

However, in many cases the additional information required to formulate these constraints is simply not available. Much of the power genome-scale methods is their potential to make predictions even in poorly characterized systems. Few metabolic flux measurements are available for most organisms, so here we have used only flux prediction techniques without flux measurements as inputs. Even in well known model organisms, regulatory interactions are only partially understood and imperfectly captured in a linear framework. In contrast, the approach we propose here uses only stoichiometric models and basic

thermodynamic constraints, which can be inferred from any annotated genome sequence [22].

#### PSEUDO: a new objective for microbial metabolism

A geometric interpretation of the PSEUDO method is presented in Figure 1D. We propose an objective function that explicitly accounts for a region of degenerate near-optimality. This region,  $\mathbf{p}$ , is bounded as in a wild-type FBA model, with the additional constraint that it includes only flux configurations with nearly optimal growth. In this case, we set a threshold of at least 90% maximal growth rate on the vectors we will consider.

$$\begin{aligned} \mathbf{b}_L &\leq \mathbf{p} \leq \mathbf{b}_U \\ \mathbf{S} \cdot \mathbf{p} &= [0] \\ P_{GROWTH} &\geq 0.90 \cdot \hat{f}_{GROWTH} \end{aligned} \quad (1)$$

The flux bounds  $\mathbf{b}_L$  and  $\mathbf{b}_U$  constrain fluxes that are known to be thermodynamically irreversible or that are

limited by media inputs. The matrix  $\mathbf{S}$  represents the biochemical stoichiometries of all metabolic reactions. The product  $\mathbf{S}\cdot\mathbf{p}$  yields the net production or consumption rate of each metabolite in the system, necessarily  $\mathbf{0}$  in the steady state. The maximum growth rate,  $\hat{f}_{GROWTH}$ , derives from a standard FBA problem. Yet compared to a single-point FBA solution, the above is a more complete and conservative expression of what we can predict with confidence regarding growth-optimal fluxes. It expresses imperfection in the proposition that metabolism works only to maximize growth rate.

We then introduce a flux vector,  $\mathbf{q}$ , representing a mutant version of the same organism. The  $\mathbf{q}$  vector is not required to show near-optimal growth, but the mutation imposes the additional bounds,  $b'_U$  and  $b'_L$ , on a subset of fluxes,  $\mathbf{q}_{MULT}$ . We propose that cellular regulation will push metabolism in the mutant, not towards a single optimal point as in MOMA, but towards a degenerate optimal region.

$$\begin{aligned} & \text{minimize : } \|\mathbf{p}-\mathbf{q}\| \\ & \text{subject to :} \\ & b_L \leq \mathbf{p} \leq b_U \qquad b_L \leq \mathbf{q} \leq b_U \qquad (2) \\ & P_{GROWTH} \geq 0.90 \cdot \hat{f}_{GROWTH} \qquad b'_L \leq \mathbf{q}_{MULT} \leq b'_U \\ & \mathbf{S}\cdot\mathbf{p} = [0] \qquad \mathbf{S}\cdot\mathbf{q} = [0] \end{aligned}$$

The above describes the geometric problem of finding the minimum distance between two polytopes:  $\mathbf{p}$  representing the region of nearly optimal growth and  $\mathbf{q}$  the space of possible fluxes limited by mutation. The point in  $\mathbf{q}$  closest to the region  $\mathbf{p}$  is the PSEUDO-predicted flux configuration for this metabolic mutant.

A PSEUDO solution will exist when the maximum mutant growth rate is less than the threshold set for the near-optimal region. Otherwise, the mutant and near-optimal flux polytopes overlap, describing a range of possible degenerate solutions. A minimum distance of 0 as a solution for (2) is immediately diagnostic for this degenerate case. A second form of degeneracy is possible in our model if the near-optimal and mutant polytopes align such that multiple solutions share the same non-zero Euclidean distance. This case is similar to classical FBA, in that an optimal objective value is defined but does not correspond to a unique solution.

As detailed in the Methods section, we can reformulate the objective  $\|\mathbf{p}-\mathbf{q}\|$  as a problem for either quadratic or conic convex programming. Therefore this formulation retains the desirable computational properties of FBA and MOMA, i.e. a guaranteed global optimum can be computed rapidly.

### PSEUDO growth predictions fall between FBA and MOMA predictions

Genome-scale optimization methods have a well-established utility for predicting growth rates of mutant

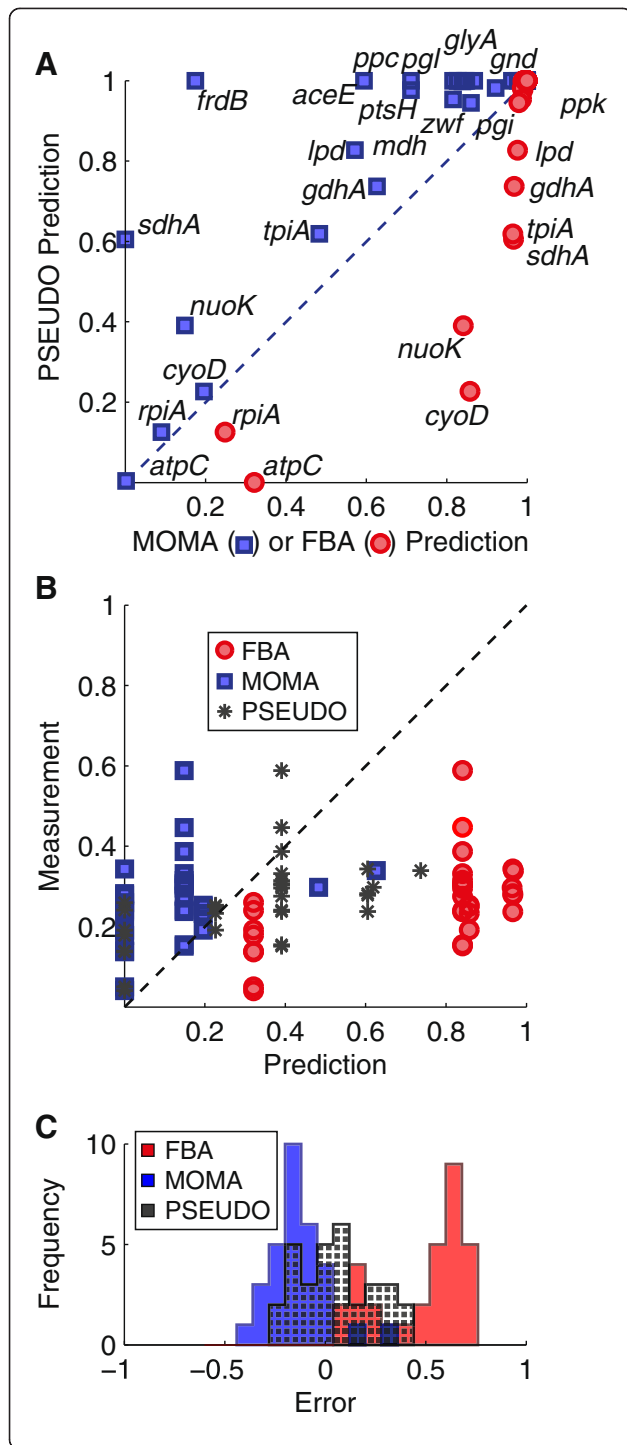
strains under a variety of conditions. We compared growth rate predictions using the PSEUDO method to predictions from the FBA and MOMA techniques, both of which are commonly used for this purpose [23]. Predictions were compared to growth rates of *E. coli* deletion mutants from the Keio collection in defined glucose medium [24]. We compiled data for 795 mutant strains that could be represented in our model and for which growth rate data was available. The results of this comparison are shown in Figure 2.

The PSEUDO method was able to correctly predict the lethality phenotype for 88% of the deletion mutants examined. Results were comparable with FBA and MOMA (88% and 87%, respectively) and consistent with previous reports [4]. A deletion was predicted lethal if the growth rate was calculated to be less than 5% of the maximum. A deletion was empirically lethal if the measured OD yield was less than 5% of the maximum reported yield. The quantitative growth rate predictions made by the three methods were in general similar, with PSEUDO growth rates within 5% of the FBA optimum for 763 of the 795 mutants examined. Overall, FBA, MOMA and PSEUDO predictions correlated equally well with yield data (Spearman's  $\rho = 0.55, 0.54, 0.55$ , respectively). This general similarity was expected. Mutations which stoichiometrically block the production of biomass must be lethal by all three methods. Mutations in pathways that are not active in these media conditions will not affect growth predictions in any model.

However, PSEUDO growth predictions were found to differ by more than 5% from either FBA or MOMA predictions for 41 of the examined mutants (Figure 2A). This subset includes mutants bearing deletions of genes involved in glycolysis, the pentose phosphate pathway and the citric acid cycle. These disruptions in central carbon processing significantly impact cell fitness and require adaptations throughout the metabolic network, providing a challenging test case for predictive theories of metabolism. We therefore examined more closely behavior of our model in these cases.

Without exception, the PSEUDO growth prediction falls between those of FBA and MOMA. This pattern highlights an important feature of the PSEUDO objective function. FBA, by definition, calculates the highest possible growth rate consistent with thermodynamics and the conservation of mass. MOMA and PSEUDO growth predictions will both fall short of this maximum if there is a trade-off between matching the optimal growth rate and matching the rest of the wild-type flux vector. PSEUDO relaxes this trade-off by considering a range of targets within the degenerate optimal region and therefore PSEUDO growth rate predictions generally exceed those of MOMA.

The growth predictions that differ among the three models are plotted against measured growth data in



**Figure 2 PSEUDO growth predictions fall between FBA and MOMA.**

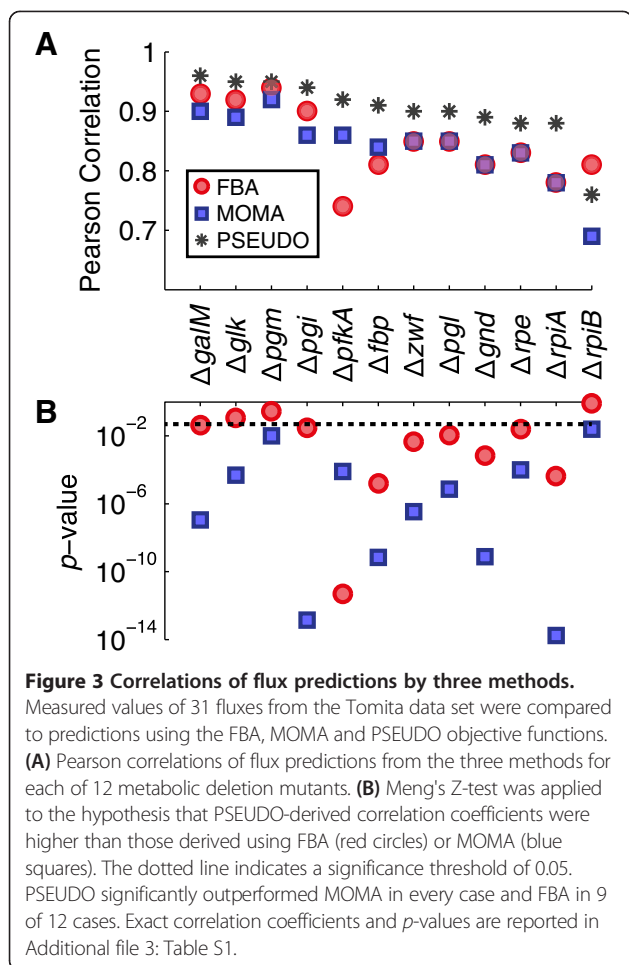
Predicted and measured yields are both normalized to a maximum of 1, corresponding to wild-type growth in these conditions. **(A)** Yields are plotted for 41 mutants for which growth predictions differ by more than 5% among the three methods. The diagonal line indicates equal predictions in PSEUDO and other methods. Note that MOMA predictions are consistently above the diagonal, and FBA consistently below. **(B)** Compared with measured growth rate data, PSEUDO growth predictions produce a rank correlation  $\rho$  of 0.63 with yield data, compared to 0.45 for the FBA method and 0.42 for MOMA. **(C)** PSEUDO predictions show no systematic bias. The error is the difference between prediction and measurement. The mean of the error distribution for PSEUDO was 0.06 with a 95% confidence interval of [0.13, -0.01] by bootstrap resampling. FBA predictions exceeded growth rates by 0.48 on average [0.39, 0.54]. The mean MOMA prediction error was -0.13 [-0.18, -0.08]. While the PSEUDO errors were unbiased, FBA and MOMA predictions were systematic over- and underestimates, respectively.

Figure 2B. The mathematical requirement that PSEUDO predictions fall between FBA and MOMA correctly recapitulates the growth phenotypes of these strains. In all cases, a comparable rank correlation was found between predicted and observed growth rates (0.45, 0.42 and 0.63 for FBA, MOMA, PSEUDO respectively). However, we found the FBA predictions to consistently overestimate growth of these strains, while MOMA predictions were an underestimate in general. Figure 2C shows a histogram of prediction errors, the difference between measured and predicted yields for each method. The average PSEUDO prediction error was 0.06, and the mean of the error distribution was not significantly different from zero by either bootstrap resampling or ANOVA statistics ( $p$ -value: 0.2). In contrast, FBA systematically overestimated growth rates by 0.48 on average ( $p$ -value:  $7.5 \cdot 10^{-9}$ ). MOMA predictions were lower than predicted values by 0.13 on average, and were significantly biased to underestimation ( $p$ -value:  $1.0 \cdot 10^{-8}$ ).

#### Precision and accuracy of PSEUDO flux predictions

We next sought to compare the biochemical flux predictions derived from our model to empirical flux measurements. The Metabolic flux rates of 31 central carbon reactions in 24 *E. coli* metabolic deletion mutants as determined by  $^{13}\text{C}$ -tracer experiments were reported by Ishii and [25]. We found that 12 of these mutants carried enzymatic deletions that could be treated within our genome-scale optimization framework, and that 27 fluxes showed measurable variation between strains. Accounting for occasional omissions in the data set, we were able to curate a total of 320 flux measurements against which to compare our predictions.

Figure 3 compares measured fluxes values to predictions derived using the FBA, MOMA or PSEUDO objective functions. To facilitate comparison, both predicted and measured fluxes were normalized to the



glucose uptake rate. Pearson correlation coefficients obtained for predictions within each of the 12 individual mutants ranged from 0.74 to 0.96. For 9 of the 12 strains, the PSEUDO method yielded a significantly higher correlation coefficient than either of the other two methods ( $p$ -value  $< 0.05$ ). In one case, standard FBA produced the best predictions. In two cases, correlations from two or more methods were statistically equivalent. Statistical significance was assessed with Meng's Z-test, which takes into account the high degree of correlation between the predictions from each method [26]. Similar significance results were obtained by bootstrap resampling. Exact correlation coefficients and  $p$ -values are reported in Additional file 3: Table S1.

For a global view of the metabolic behavior predicted by the FBA, MOMA and PSEUDO objective functions, we aggregated and examined predictions for all 320 measured fluxes across 12 mutants (Figure 4ABC). Across all 320 fluxes, PSEUDO was more predictive than FBA and MOMA (Pearson correlation coefficients of 0.86, 0.84 and 0.91 respectively for FBA, MOMA, and PSEUDO;  $p$ -value:  $2.4 \cdot 10^{-12}$ , Meng's Z-test). Because measured flux values span several orders of magnitude,

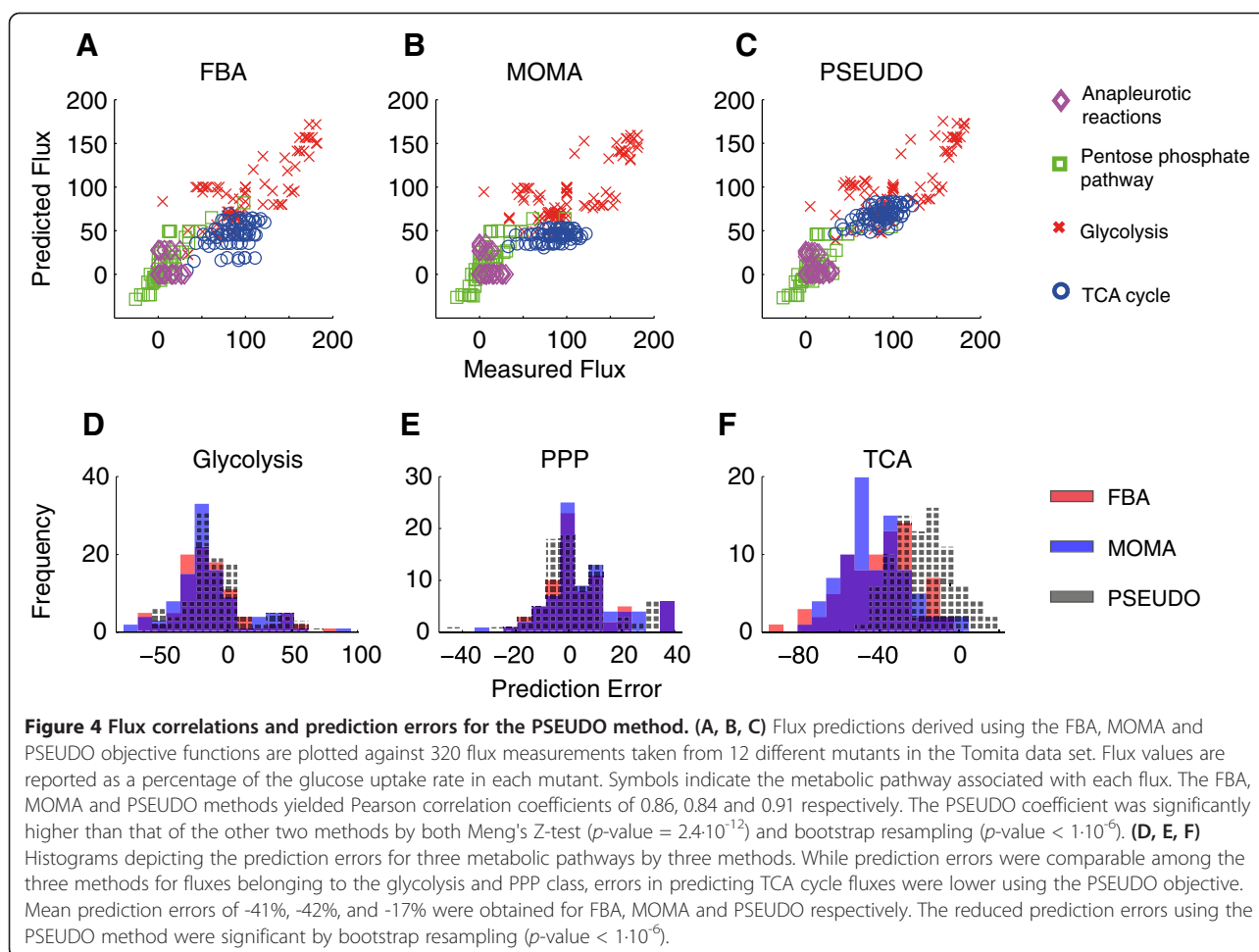
we also compared predictions to data using rank correlation coefficients that are less influenced by numerical outliers. An overall Spearman rank correlation of 0.82, 0.80, and 0.87 was obtained with FBA, MOMA and PSEUDO respectively. The higher coefficient from PSEUDO was again significant by bootstrap resampling ( $p$ -value  $< 1 \cdot 10^{-6}$ ).

Prediction accuracy was found to vary substantially for different pathways within central carbon metabolism. The histograms in Figure 4DEF compare prediction errors by each method in reactions belonging to glycolysis, the pentose phosphate cycle (PPP) and the tricarboxylic acid cycle (TCA) reactions. While prediction errors within glycolysis and the PPP were comparable among the 3 methods, the PSEUDO method produced significantly lower prediction errors within the TCA cycle. The FBA, MOMA and PSEUDO methods produced mean prediction errors of -41%, -42%, and -17% respectively. Flux variation within this class was also better predicted by PSEUDO. For TCA cycle reactions, FBA, MOMA and PSEUDO yielded Pearson correlations of 0.41, 0.38 and 0.60 respectively ( $p$ -value: 0.01). Thus, both the absolute level of TCA cycle flux and flux variations within the TCA cycle were better predicted by PSEUDO. Relatively high error in TCA cycle predictions when using growth as an objective in carbon-limited conditions has been reported in other models [15,27]. We found predictions within the TCA cycle to be both the most error-prone and the most revised under our model, being responsible for most of the improved performance. We found no other significant differences in predictions from the 3 methods in glycolysis, the PPP, anapleurotic, or secretion reactions.

### Sensitivity analysis of TCA cycle predictions

To understand why the PSEUDO objective function improves flux predictions for the TCA cycle, we chose to investigate central carbon metabolism in greater detail using the *zwf* mutant as a case study (Figure 5). Carbon consumed in the form of glucose may be converted to biomass, fully oxidized to  $\text{CO}_2$  or secreted as reduced organic metabolites. We reasoned that the metabolic strategy used to fulfill a given objective function would be reflected in the way carbon is partitioned among these three final forms.

We observed significant differences in the ultimate fate of carbon in the *zwf* mutant as predicted with the FBA, MOMA and PSEUDO objective functions (Figure 5A). The FBA model predicts the largest flux of carbon into biomass (63%, 48%, 56% and for FBA, MOMA, PSEUDO, respectively). This is consistent with the mathematical requirement that FBA identify the highest possible growth rate. The MOMA model predicts a significant amount of carbon secreted in the form of organic metabolites (0%,



20%, and 2% for FBA, MOMA, PSEUDO, respectively). The profile of secreted metabolites also varies among the models. While PSEUDO predicts mainly acetate secretion, MOMA predicts significant secretion of 33 diverse metabolites including urea, glutamate, leucine and tryptophan. Organic secretion is likely driven by the MOMA objective to match the WT flux profile. In effect, the MOMA attempts to produce metabolites at WT levels that cannot be completely consumed and partially secretes the difference.

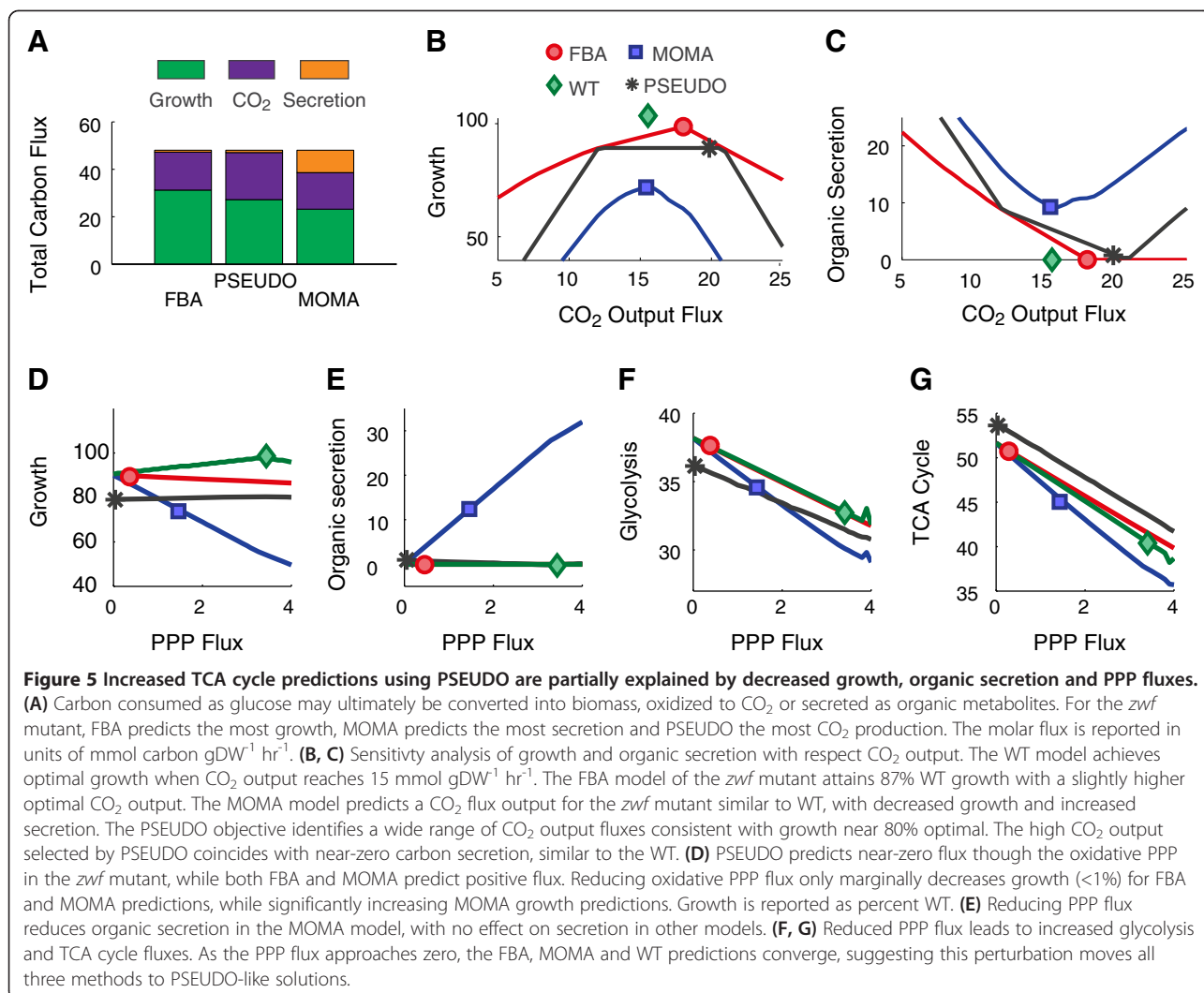
The PSEUDO objective function predicts the highest CO<sub>2</sub> output (37%, 32% and 42% for FBA, MOMA, PSEUDO respectively). This increase in total carbon oxidation is consistent with the higher TCA cycle flux predicted by PSEUDO. It may be explained by the relaxation in PSEUDO of strict optimality requirements used in the other methods. PSEUDO is not explicitly driven, like FBA, to optimize biomass production. Nor does it produce excess metabolites and secrete them, like MOMA. With less carbon flux dedicated to these objectives, the PSEUDO model retains more carbon to fully oxidize.

We next performed a sensitivity analysis to characterize the effect of CO<sub>2</sub> output on the behavior of each

model. We constrained the total CO<sub>2</sub> output flux to a series of specific values near the WT optimum and re-solved the *zwf* mutant model to determine the fluxes to growth and organic secretion. The resulting plots reveal the trade-offs confronted in each objective function when assigning central metabolic fluxes near optimality (Figure 5BC).

In the FBA model, the *zwf* mutant reaches a maximum growth rate of 87% WT with a slightly increased CO<sub>2</sub> output. The secretion flux approaches zero as growth attains a maximum, indicating that secreting organic carbon is costly to growth. The MOMA model predicts less growth and more secretion, but also exhibits a trade-off between the two. Any deviation of CO<sub>2</sub> output levels from the WT optimum results in less growth and more secretion. In contrast, the PSEUDO solution is not found near a local growth maximum. Instead, the PSEUDO method identifies a range of CO<sub>2</sub> output fluxes that are consistent with near-optimal growth. In the absence of a trade-off with growth, the PSEUDO objective is free to match other features of the WT flux vector. In this case, it selects a high CO<sub>2</sub> flux that coincides with low secretion, similar to the WT.





To further characterize the assignment of TCA cycle fluxes in PSEUDO, we extended our analysis to examine the CO<sub>2</sub> produced by the three main carbon-oxidizing pathways: Glycolysis, the PPP and the TCA cycle (Figure 5EFG). We observed a striking qualitative difference in the behavior of PSEUDO and the other objective functions. Uniquely, the PSEUDO objective function predicts that no CO<sub>2</sub> is generated through the PPP. In contrast, FBA and MOMA predict that significant CO<sub>2</sub> production through the oxidative reactions of the PPP (2.5% and 10% of the total for FBA and MOMA, respectively).

In fact, the *zwf* mutant shows no oxidative PPP activity in published observations [28]. During normal glucose-limited growth, the *zwf* gene product (together with *pgl*) supplies 6-phosphogluconate to the oxidative reactions of the PPP. In the *zwf* mutant, the FBA and MOMA models alternately supply this molecule through the action of glucose dehydrogenase and glucokinase. While *E. coli* is capable of oxidizing glucose directly to gluconate, the activity

appears only under a narrow range of conditions not found in laboratory cultures [29].

As revealed by sensitivity analysis, small alterations in the PPP flux had significant consequences for the predictive power of each model. Constraining the oxidative PPP flux to zero only slightly decreased the growth predictions for the FBA WT and *zwf* models (Figure 5D). This indicates that a zero-PPP solution exists near the WT optimum. However, neither FBA nor MOMA identified this solution. In the case of FBA, glucokinase activity bestows a small growth advantage, while MOMA is driven to match the high PPP flux of the WT.

When PPP flux was decreased, the MOMA model predicted significantly higher growth and lower secretion, indicating that a zero-PPP solution alleviated the trade-off between these two fluxes in MOMA. As PPP fluxes approach zero, glycolytic and TCA cycle fluxes increase and converge in all models, indicating that similar solutions were found by FBA, MOMA and PSEUDO under

reduced PPP flux. We conclude that the low TCA cycle predictions in the MOMA and, to a lesser extent, the FBA model are partially a consequence of positive oxidative PPP fluxes. In contrast, PSEUDO approaches a solution with near-zero PPP flux, still consistent with near-optimal WT growth, and much less disruptive for the *zwf* metabolic network.

In summary, several features of the PSEUDO objective appear to contribute to higher and more accurate predictions of TCA cycle fluxes for the *zwf* mutant. The PSEUDO model grows less than FBA and secretes less than MOMA, leaving more carbon available to oxidize. PSEUDO is less constrained in matching growth rates, predicting equal growth from a wide range of possible CO<sub>2</sub> output fluxes. Within this range, the PSEUDO objective selects a flux profile that matches key features of the WT. High CO<sub>2</sub> and TCA fluxes are consistent with glycolytic fluxes and a secretion profile that resemble the WT.

In the case of the *zwf* mutant, the qualitative features of the PSEUDO prediction including the high growth rate, the increased CO<sub>2</sub> production and zero oxidative PPP flux agree with published phenotypic observations [28]. In contrast, the FBA and MOMA predictions fail to reproduce one or more of these three qualitative phenotypic features.

#### **PSEUDO predictions are robust to the near-optimal growth threshold**

We next sought to characterize the global behavior of the PSEUDO model as a function of the threshold that we use to define near-optimal growth. This value, set to 90% optimal growth in Equation 2, is a required input parameter for our model that has no equivalent in either FBA or MOMA. Mathematically, this represents a degree of uncertainty in the hypothesis that metabolism optimizes only growth rate. Biologically, this threshold could be understood as the point at which selection for growth is counterbalanced by other, unknown, selection pressures or by inherent noise. The results presented above were made using a threshold parameter of 90% optimal growth. This selection was guided by growth variability reported in the literature. For example, the reported growth rates of metabolic mutants of *B. subtilis* suggest that metabolism may be sub-optimal for growth at roughly this level under laboratory conditions [13].

We found that PSEUDO predictions were remarkably stable as the near-optimal growth threshold was varied from 80-99%, as shown in Figure 6. Both Pearson and Spearman correlation values for PSEUDO predictions reached a maximum with the growth threshold set to 90%, and declined as near-optimal growth converged to maximum theoretical growth (Figure 6AB). We observed no qualitative differences in model behavior across this

parameter range (Figure 6CDEF). This behavior is consistent with the convex shape of flux space. In a convex space, variability tends to increase rapidly for small deviations from optimality, then decelerate and plateau at moderate deviations [30-32]. Robustness with respect to the selected threshold is an important feature of the PSEUDO model, as this parameter may be difficult to measure in practice.

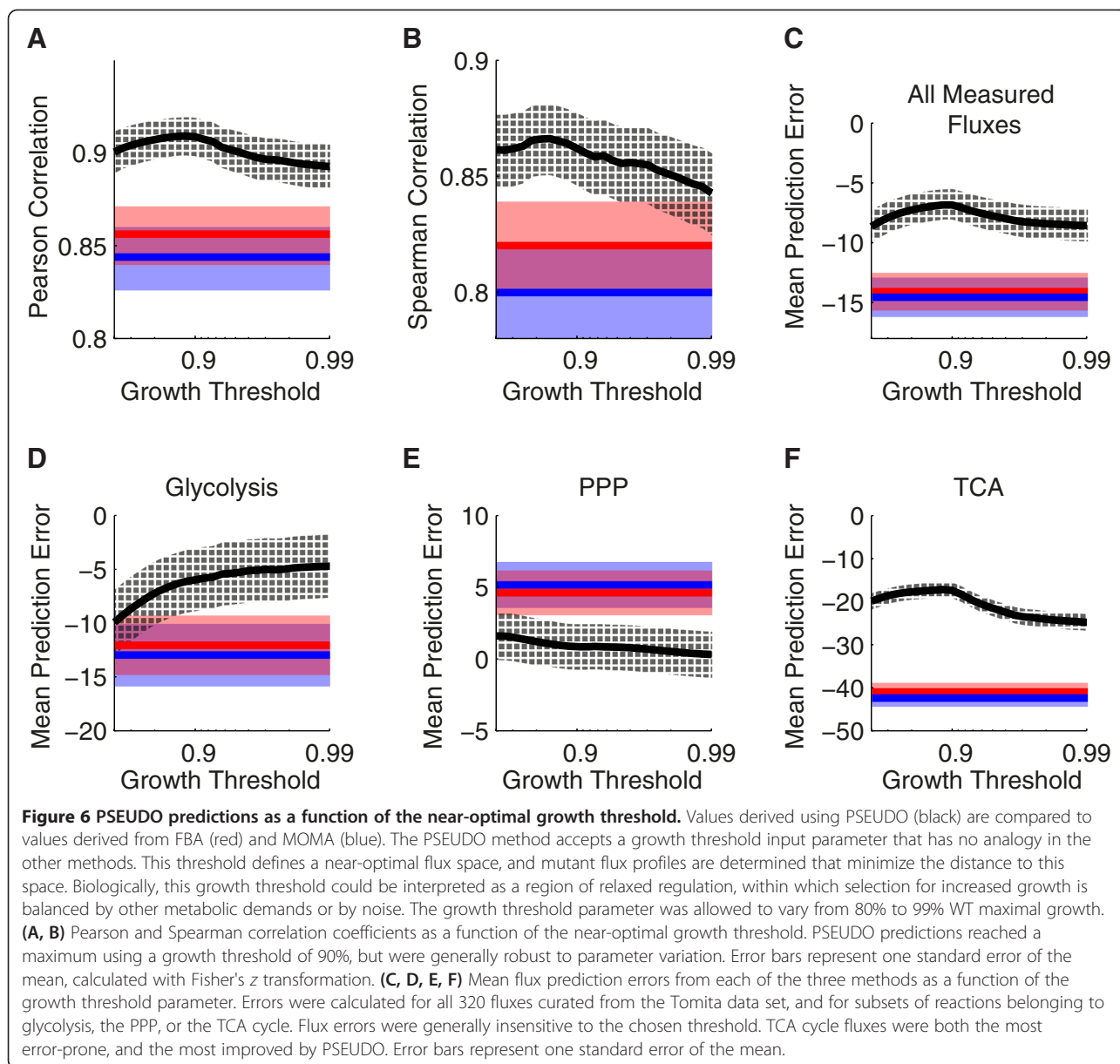
For very high or low values of the growth threshold, the practical application and biological interpretation of the PSEUDO model becomes more difficult. In practice, we were unable to compute solutions with threshold values higher than 99.9%, as extremely narrow range constraints on individual fluxes are known to challenge interior-point optimization solvers [33]. In this regime, we expect that PSEUDO predictions will become MOMA-like as the near-optimal region shrinks to become the optimal region. For low values of the growth threshold, the near-optimal region grows eventually to include the MOMA and FBA solutions and specific PSEUDO predictions are undefined.

#### **A cloud theory of metabolic regulation**

The PSEUDO formulation includes a limit on the power of growth rate alone to determine metabolic behavior. This may represent simply a formal accounting for uncertainty, which is neglected in simpler models. Alternatively, our objective function may be interpreted as expressing an organizational principal at work in metabolism. The metabolic network may not be perfectly adaptable in the service of growth, as imagined in FBA, nor strictly committed to a singular flux profile, as postulated by MOMA. The PSEUDO method mediates between these perspectives, attributing to metabolism an intermediate level of flexibility. Our model suggests that regulation drives metabolic fluxes to a certain range of values, but that fluctuations within that range are fitness-neutral and unregulated. On this hypothesis, we expect that regulation will allow fluxes to vary in proportion to the size of their near-optimal range.

To compare theoretical and observed flux variability, we first sought a properly normalized measure of the size of the degenerate optimal polytope in each dimension. While it is computationally infeasible to describe this space completely, [12,34], it is possible to estimate its shape probabilistically [32]. As described in the methods, we used a Monte Carlo sampling technique to generate a set of 3000 random points uniformly distributed within the degenerate optimal region [35,36]. These points allow an unbiased estimate of the distribution of values a given flux may attain without compromising near-optimal growth.

An ideal measure of flux variability *in vivo* would compile flux values from individual wild-type cells. Unfortunately,



no such single-cell resolution dataset is currently available. Instead, we examined flux variability at the population level under two sources of genetic perturbation. The Tomita data set includes 31 fluxes measured in chemostat cultures of 24 deletion mutants for enzymes in central carbon processing [25]. Viable glycolysis and pentose phosphate pathway deletion mutants often exhibit substantially revised flux profiles and test the limits of metabolic plasticity. The Sauer data set measures 24 central carbon fluxes in 91 transcription factor deletion mutants [37]. These regulatory disruptions globally alter flux profiles while leaving the enzymatic network intact. We reasoned that these data sets together would allow us to characterize the tendency of individual fluxes to vary under perturbation.

Figure 7 compares our measures of theoretical and observed variation. For each flux in our data set, we compared the coefficient of variation (CV) derived from computational Monte Carlo sampling (Figure 7A) to the CV from published measurements (Figure 7B). Measured variability in both data sets was well matched by theoretical variability within the degenerate optimal region (Figure 7CD). The Sauer flux measurements produced a rank correlation of 0.72 (p-value:  $6.7 \cdot 10^{-5}$ ). For the Tomita data, Spearman's  $\rho$  was 0.87, (p-value:  $5.6 \cdot 10^{-9}$ ). Exact values for the estimated and measured variability of each flux are reported in Additional file 4: Table S2. The observed correlation between predicted degeneracy and measured variation supports a model in which metabolism

may adopt many possible flux configurations without compromising growth rate.

### Conclusions

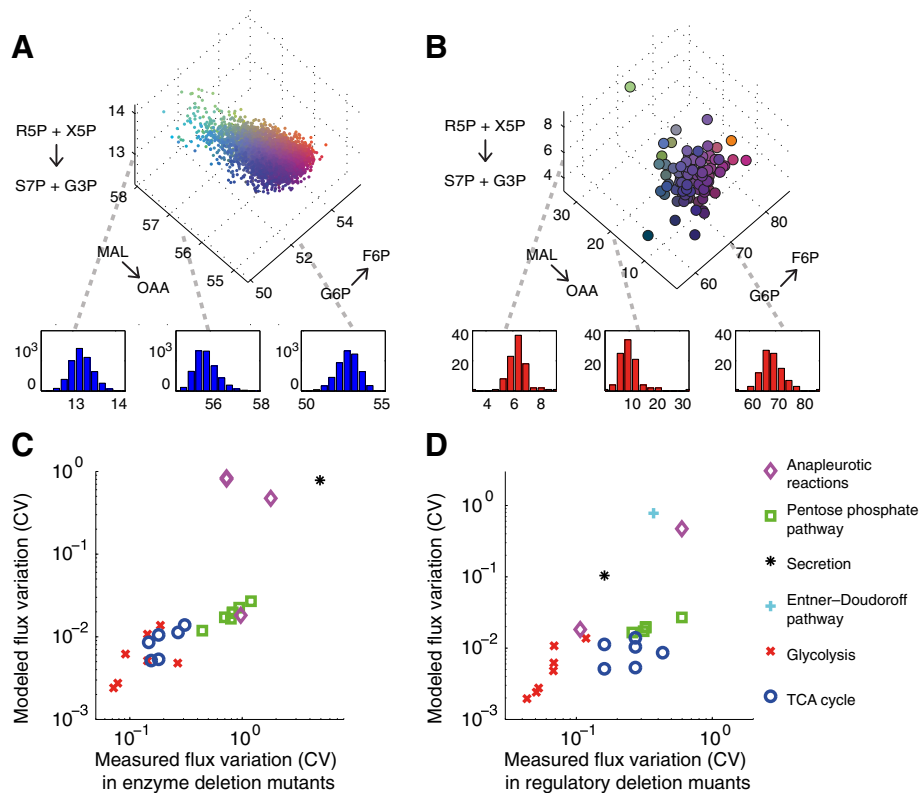
If the global behavior of metabolic regulation can be expressed mathematically as an objective function, then an optimization algorithm can identify the region of flux space that achieves this objective. The remarkable success of FBA, MOMA and other approaches in estimating metabolic behavior raises the tempting possibility that a truly general objective function may yet be found. To that end, recent work seeks to systematically compare the predictive performance of candidate metabolic objectives under diverse conditions [27,38,39].

Yet a single objective function may never fully capture the competing demands on a living biochemical network. A more versatile model of metabolism might be expressed in terms of multiple objectives and a formal

framework for describing the tradeoffs among them [30,40,41]. However, combined objectives have limited predictive power without prior knowledge of how they will be balanced in a particular organism.

An alternative approach is to treat suboptimality as an inherent and irreducible feature of biology. Deviations from optimality may themselves be well ordered, and can reveal the action of novel selective forces and physiological constraints [2,42,43]. We have shown that variations in metabolic fluxes under perturbation can be largely explained if variation is constrained to be not growth-neutral, but nearly so. In general, metabolic systems biologists will be challenged to distinguish model errors due to a flawed objective function from variation which is suboptimal but constrained, or variation which is stochastic and does not satisfy any objective.

Finally, metabolism may simply be noisy. Metabolic flux space is large and high dimensional. It could remain so,



**Figure 7 A cloud theory for metabolic regulation.** The PSEUDO method hypothesizes the existence of a degenerate optimal region. Fluxes are regulated to approach this region, but allowed to vary freely within it. **(A)** Variability of fluxes within the degenerate optimal region. As described in the methods, we used Markov chain Monte Carlo sampling to produce 3000 randomly distributed fluxes, each consistent with at least 99% maximal growth. 3 specific fluxes are plotted on these axes, colored to emphasize their value in each dimension. Red, green and blue content corresponds to value in x, y and z respectively. Histograms present overall distributions of individual fluxes. **(B)** Variation in measured values for 24 fluxes across 91 transcription factor deletion mutants from the Sauer data set. Measured data is plotted similarly to the Monte Carlo generated data. **(C, D)** Computationally estimated flux variability within the degenerate optimal region correlates well with measured variability in both the Sauer **(C)** and Tomita **(D)** data sets. The coefficient of variation (CV) of the distributions is a normalized expression of both measured and predicted flux variability. Marker shapes are used to indicate the metabolic class of each flux. Metabolic fluxes that are only weakly coupled to biomass production can vary widely within the degenerate optimal region. The same fluxes are more likely to vary in under perturbation in published data sets.

even after all physical constraints have been satisfied and all functional objectives have been met. This would leave ample room for variation in flux rates between cells and within a cell over time. The metabolic behavior of a population may not be adequately described by any single flux profile. Seen from this perspective, suboptimal and degenerate solutions are not merely a technical inconvenience leading to non-unique predictions. They are rather an inherent and evolved property of metabolic organization.

We propose that a degenerate and variable biochemical flux profile could support a more robust cellular metabolism. Robustness, the capacity to perform an essential function despite perturbation, is a common property of biological systems [44]. Metabolic networks are already known to be replete with structural features that support robust function.

For example, the genes that code for many essential enzymes are carried in duplicate. Isoenzymes have the opportunity to diverge and specialize through evolution, but often remain capable of substituting for each other in the event that one becomes mutated or otherwise compromised [45,46]. This redundancy can extend from single genes to whole pathways, like the glyoxylate shunt, the methylglyoxal bypass, or the Entner–Doudoroff pathway, that provide alternate routes to key metabolic precursors [47].

Further, metabolic fluxes are typically robust to fluctuations in the abundance of the enzymes that catalyze them. While pathway fluxes can be controlled by the coordinated regulation of multiple enzymes, the total flux through a metabolic pathway rarely depends strongly on variations in the expression any single enzyme [48]. Most enzymes are produced in significant excess to the needs of a cell in standard conditions [49]. Perturbations that alter enzyme production are therefore unlikely to result in any given enzyme becoming flux-limiting. These sources of robustness may be particularly important given ubiquitous and substantial stochastic variation in protein copy number [50].

Degenerate optimality of the sort we have described would allow another level of functional redundancy. Even confronted with perturbations in enzyme function, regulation or copy number that destabilize intracellular fluxes, microbial growth would be unaffected so long as fluxes remain within the degenerate optimal region. Every cell must confront fundamental physical limits on its ability to regulate the internal environment [14]. A degenerate organization of metabolic fluxes would sidestep these limits, uncoupling growth rate from strict regulation which may be energetically costly or mechanically impossible.

If selection and therefore regulation were relaxed with respect to growth rate, opportunities would emerge for selection to act on other features of the metabolic flux

vector. For example, cells could anticipate a changing environment by minimizing the need to redirect fluxes under different metabolic conditions that are likely to appear [30]. Phenotypes within this open space could in principle be selected for any biochemical feature required by the local environment, without sacrificing near-optimal growth. Two bacteria may occupy the same metabolic niche and grow at the same rate, yet exhibit differently specialized flux profiles.

Strategic metabolic heterogeneity at the population level is already known in some systems. For example, *E. coli* populations under certain conditions display a bimodal distribution for expression of the lactose permease [51]. Individual cells can switch randomly between high or low enzyme levels, then propagate that state for a time scale of generations. A subset of the population is able to rapidly take up lactose, should it become available, while other subsets remain specialized for other carbon sources. This may be a way of hedging bets in unpredictable nutritive environments, without requiring each cell to prepare for all possible conditions [52,53]. Degenerate and growth-neutral variation in intracellular fluxes would allow another opportunity for adaptive bet hedging, as certain flux configurations may tend to better anticipate environmental changes. By tolerating growth-neutral variability, a more relaxed metabolic regulation would expand the range of accessible individual phenotypes.

Under this cloud theory of metabolic regulation, we expect flux magnitudes to exhibit substantial diversity at the single-cell level. This variation should occur even in clonal populations and uniform environments. It should correlate with the dimensions of the degenerate optimal flux region and have very little effect on growth rate. Phenotypic heterogeneity is increasingly seen to underlie complex emergent behaviors in genetic competence [54], antibiotic resistance [55], lysogeny [56], apoptosis [57], and stem cell induction [58]. Many other fundamental cellular properties are now known to vary substantially within populations [59-61]. We expect that the stochastic variations of metabolic fluxes within single cells, once measured, will reveal new modes of robust biological order.

## Methods

### FBA and MOMA

Standard FBA solves for a vector of metabolic fluxes,  $\hat{f}$ , that maximizes cellular growth rate,  $\mathbf{f}_{GROWTH}$ .

$$\begin{aligned} &\text{maximize : } \mathbf{f}_{GROWTH} \\ &\text{subject to :} \\ &\quad b_L \leq f \leq b_U \\ &\quad S \cdot f = [0] \end{aligned} \tag{3}$$

The growth flux  $\mathbf{f}_{GROWTH}$  represents metabolites leaving the system in the form of new biomass with a

composition determined by measurement [62]. The matrix  $\mathbf{S}$  represents the biochemical stoichiometries of all metabolic reactions. The product  $\mathbf{S}\cdot\mathbf{f}$  yields the net production or consumption rate of each metabolite in the system, necessarily  $\mathbf{0}$  in the steady state. The flux bounds  $\mathbf{b}_L$  and  $\mathbf{b}_U$  constrain fluxes that are known to be thermodynamically irreversible or that are limited by media inputs. Mutations in FBA are modeled by setting the appropriate flux bounds to zero.

The MOMA method [15] identifies a flux vector,  $\mathbf{m}$ , with minimum Euclidean distance to an optimal wild-type solution,  $\hat{\mathbf{f}}$ , subject to the constraints of mutation.

$$\begin{aligned} & \text{minimize : } \|\mathbf{m}-\hat{\mathbf{f}}\| \\ & \text{subject to :} \\ & \quad \mathbf{b}_L \leq \mathbf{m} \leq \mathbf{b}_U \\ & \quad \mathbf{b}'_L \leq \mathbf{m}_{MUT} \leq \mathbf{b}'_U \\ & \quad \mathbf{S}\cdot\mathbf{m} = [\mathbf{0}] \end{aligned} \quad (4)$$

The stoichiometric matrix,  $\mathbf{S}$ , and the bounds,  $\mathbf{b}$ , are defined as in (3). The modified bounds  $\mathbf{b}'_L$  and  $\mathbf{b}'_U$  are applied to the subset of fluxes eliminated by mutation,  $\mathbf{m}_{MUT}$ .

#### Quadratic and conic formulations of PSEUDO

The PSEUDO method finds the minimum Euclidean distance between two high-dimensional flux polytopes,  $\mathbf{p}$  and  $\mathbf{q}$ . Constraints are added to place  $\mathbf{p}$  within the region of degenerate optimality and impose a mutation on  $\mathbf{q}$ .

$$\begin{aligned} & \text{minimize : } \|\mathbf{p}-\mathbf{q}\| \\ & \text{subject to :} \\ & \quad \mathbf{b}_L \leq \mathbf{p} \leq \mathbf{b}_U \quad \mathbf{b}_L \leq \mathbf{q} \leq \mathbf{b}_U \\ & \quad P_{GROWTH} \geq 0.9 \cdot \hat{\mathbf{f}}_{GROWTH} \quad \mathbf{b}'_L \leq \mathbf{q}_{MUT} \leq \mathbf{b}'_U \\ & \quad \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} = [\mathbf{0} \quad \mathbf{0}] \end{aligned} \quad (5)$$

Note that (5) is identical to (2) from the main text, except we have concatenated the conservation-of-mass constraints  $\mathbf{S}\cdot\mathbf{p} = \mathbf{0}$  and  $\mathbf{S}\cdot\mathbf{q} = \mathbf{0}$ . This form is standard for most solver algorithms.

In order to be analyzed with the powerful tools of convex programming, an optimization problem must be formulable with a convex objective under convex constraints [63]. The system of constraints above is composed only of linear inequalities, and is therefore convex. It remains only to provide a general convex formulation of the objective function:  $\|\mathbf{p}-\mathbf{q}\|$ . We first substitute the definition of a Euclidean distance between two vectors.

$$\text{minimize : } \sqrt{\sum_i (\mathbf{p}_i - \mathbf{q}_i)^2} \quad (6)$$

Minimizing a positive radicand is equivalent to minimizing the root, so we neglect the radical.

The objective then expands to:

$$\text{minimize : } \sum_i \mathbf{p}_i^2 - 2\mathbf{p}_i\mathbf{q}_i - \mathbf{q}_i^2 \quad (7)$$

Finally we reformulate the above as a linear algebraic expression, with  $\mathbf{I}$  representing the identity matrix:

$$\text{minimize : } \begin{bmatrix} \mathbf{p} & \mathbf{q} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{I} & -\mathbf{I} \\ -\mathbf{I} & \mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \end{bmatrix} \quad (8)$$

The square matrix in the above expression is symmetric and positive semidefinite, with eigenvalues of only 2 or 0. This is sufficient to guarantee convexity and polynomial solvability [64].

It is also possible to formulate PSEUDO as conic optimization problem. Conic optimization is among the most general forms of nonlinear convex optimization treated by commercial solvers. In this case we add a vector  $\mathbf{x}$  to our variable set with the linear constraint that  $x_i = \mathbf{p}_i - \mathbf{q}_i$ . The  $\mathbf{x}$  vector and a new variable,  $z_{DIST}$ , are further constrained to lie within a convex quadratic cone  $C$  such that  $z_{DIST}$  is at least the Euclidean distance between  $\mathbf{p}$  and  $\mathbf{q}$ . Minimizing  $z_{DIST}$  under these constraints produces the shortest distance from  $\mathbf{p}$  to  $\mathbf{q}$ .

$$\begin{aligned} & \text{minimize : } z_{DIST} \\ & \text{subject to :} \\ & \quad \mathbf{b}_L \leq \mathbf{p} \leq \mathbf{b}_U \quad \mathbf{b}_L \leq \mathbf{q} \leq \mathbf{b}_U \\ & \quad P_{GROWTH} \geq 0.99 \cdot \hat{\mathbf{f}}_{GROWTH} \quad \mathbf{b}'_L \leq \mathbf{q}_{MUT} \leq \mathbf{b}'_U \\ & \quad \begin{bmatrix} \mathbf{S} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} & \mathbf{0} \\ \mathbf{I} & -\mathbf{I} & -\mathbf{I} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{p} \\ \mathbf{q} \\ \mathbf{x} \end{bmatrix} = [\mathbf{0} \quad \mathbf{0} \quad \mathbf{0}] \\ & \quad [\mathbf{x} \quad z_{DIST}] \in C \quad := \left\{ z_{DIST} \geq \sqrt{\sum_i x_i^2} \right\} \end{aligned} \quad (9)$$

#### Linear, quadratic, and convex programming

All convex programming was implemented in MATLAB and solved using MOSEK optimization software.

All *E. coli* metabolic simulations were implemented using the iAF1260 model, incorporating stoichiometric and thermodynamic but not regulatory constraints [4]. Media parameters were defined minimal glucose media, using standards set by the model authors. Glucose availability was set to 8 mmol per gram dry weight (gDw) per hour, oxygen to 18.5 mmol gDw<sup>-1</sup> hr<sup>-1</sup>, with ammonia, phosphate, sulfate and trace minerals available in excess. Growth-associated ATP consumption was set to 59.81 mmol gDw<sup>-1</sup>, with 8.39 mmol ATP gDw<sup>-1</sup> hr<sup>-1</sup> required for growth-independent cellular maintenance. Model conditions correspond to a standard media formulation as used in the experimental data set [25]. Flux values were normalized to the glucose uptake rate

of each mutant, and are reported as a percentage of this value.

The MOMA solver requires a wild-type flux distribution as input. In all cases, this was calculated as an FBA prediction for the wild type strain with the same media conditions and biomass composition. Empirically measured fluxes were not an input to any solver.

A secondary objective function was used with the FBA and MOMA techniques to minimize the 1-norm of a growth-optimal solution vector. For the FBA method, this secondary objective was applied directly to the mutant solution. For MOMA, the secondary optimization was applied to the FBA-derived wild-type flux vector.

Unless otherwise indicated, the PSEUDO solutions reported were calculated using a growth threshold parameter of 90% the WT maximum, as calculated with standard FBA.

Sensitivity analysis was performed by constraining the fluxes corresponding to CO<sub>2</sub> output or phosphogluconate dehydrogenase to a range of exact values and re-solving each model. Molar carbon secretion fluxes were obtained as the sum of all output fluxes weighted by the number of carbon atoms in each output molecule.

#### Uniform sampling from flux space

Random flux configurations were generated by Monte Carlo sampling [35]. We used an artificially centered hit-and-run algorithm to produce a series of points within the degenerate optimal flux polytope [36]. As this series is extended, it asymptotically approaches a uniform random distribution.

Briefly, a random set of initial points was generated within the degenerate optimal region by solving a linear programming objective to maximize a random vector chosen on the unit sphere. These solutions sample the extreme vertices. The hit-and-run algorithm is initialized from the center of these extrema, which by the definition of convexity must also lie within the degenerate optimal region. A direction is chosen as the difference between a random extreme point and the center point. Then a line is drawn in this direction from the initial point to the region boundary and a new point is selected randomly from this line. After a sufficiently large number of such movements, the point reached is as drawn from a uniform distribution.

Following Bordbar *et al.*, we define a mixed fraction for a set of points that allowed us to determine when our algorithm has reached an approximately uniform distribution [35]. We partition the initial points into two sets along the median value of any flux. The mixed fraction is defined as the number of points which cross the line as the algorithm progresses. The mixed fraction approaches 50% asymptotically, with the algorithm terminated at a value of 53%.

## Additional files

**Additional file 1: Figure S1.** Degeneracy in Metabolic Flux Analysis. (A, B) Two alternate flux distributions in central carbon metabolism that support equally optimal growth. Configuration A uses the pentose phosphate cycle to oxidize glucose. Configuration B runs the pentose phosphate cycle in reverse only as a source of precursor metabolites. While A produces more NADPH, B produces more pyruvate for the TCA cycle. Both options identically supply the ATP, reducing equivalents and carbon skeletons needed for growth.

**Additional file 2: Figure S2.** Fluxes vary widely within the degenerate optimal range. As described in the main text, the degenerate range is the gap between the minimum and maximum possible values for each flux that support above a threshold near the maximum value. (A, B, C) The degenerate range in these three panels is calculated under the constraint of 99.9%, 99% and 95% maximal growth, respectively. The flux magnitude is calculated for the wild type system using FBA. Fluxes above the grey dotted line can vary by more than 100 times their predicted value without compromising growth. Fluxes on the red dotted line can vary by exactly 1%. Infinite flux variability is possible for fluxes in futile cycles. Zero flux variability indicates a metabolic reaction that cannot occur under these media conditions. Example enzymes from various metabolic pathways are indicated. (D) Flux variability for selected reaction classes. For nonzero fluxes, fold degeneracy is the degenerate range divided by the FBA-predicted flux magnitude. The median fold degeneracy is plotted for each class, with bars indicating the 10th and 90th quantiles.

**Additional file 3: Table S1.** Correlations of flux predictions by three methods. 31 Measured flux values from the Tomita data set were compared to predictions using the FBA, MOMA and PSEUDO objective functions. Reported values are Pearson correlation coefficients. Meng's Z-test was used to test the hypothesis that the PSEUDO-derived correlations were higher than those from each other method. Similar significance results were obtained by bootstrap resampling. Entries in bold and marked with an asterisk indicate that the PSEUDO method was not more predictive than both other methods at a  $p$ -value less than 0.05.

**Additional file 4: Table S2.** Flux variability under perturbation and the shape of the degenerate optimal region. As described in the main text, flux measurements were collected from two published data sets. The Tomita data set consists of 31 fluxes measured in 24 deletion mutants for central carbon processing enzymes. The Sauer data set includes 24 fluxes from 91 deletion mutants for regulatory transcription factors. The standard deviation,  $\sigma$ , divided by the mean flux value yields the coefficient of variation, CV, a unitless measure of flux variability. The Markov Chain Monte Carlo method was used to generate 3000 random points uniformly distributed within the degenerate optimal region of the iAF1260 FBA model of *E. coli*. Each of these points corresponds to a possible flux profile sustaining growth at at least 90% of the theoretical maximum. The mean and standard deviation of the MCMC-generated points yield an estimated CV representing the variation possible for each flux while still maintaining near-optimal growth. As depicted in Figure 7 of the main text, simulated flux variability correlates with measured variability from both data sets. This suggests a model in which relatively tight regulation is applied only to fluxes that strongly impact cell growth. Inversely, substantial variation is tolerated in fluxes with less impact on growth rate.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

EW and TL constructed the theory and derived the model predictions. EW, TL and PS analyzed the results and drafted the manuscript. All authors read and approved the final manuscript.

#### Acknowledgements

We thank Daniel Segrè, Mike Springer, Tim Mitchison, Andrew Murray and Patrick Boyle for discussions and manuscript revisions.

Received: 3 June 2013 Accepted: 30 September 2013  
Published: 3 October 2013

## References

- Joyce AR, Palsson BØ: Toward whole cell modeling and simulation: comprehensive functional genomics through the constraint-based approach. *Prog Drug Res* 2007, **64**(265):267–309.
- Parker GA, Smith JM: Optimality theory in evolutionary biology. *Nature* 1990, **348**:27–33.
- Palsson B: *Systems biology: properties of reconstructed networks*, Illustrated, reprint. Cambridge; New York: Cambridge University Press; 2006.
- Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ: A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007, **3**:121.
- Beard DA, Babson E, Curtis E, Qian H: Thermodynamic constraints for biochemical networks. *J Theor Biol* 2004, **228**:327–333.
- Covert MW, Palsson BØ: Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J Theor Biol* 2003, **221**:309–325.
- Zomorodi AR, Suthers PF, Ranganathan S, Maranas CD: Mathematical optimization applications in metabolic networks. *Metab Eng* 2012, **14**:672–686.
- Segrè D, DeLuna A, Church GM, Kishony R: Modular epistasis in yeast metabolism. *Nat Genet* 2004, **37**:77–83.
- Motter AE, Gulbahce N, Almaas E, Barabási AL: Predicting synthetic rescues in metabolic networks. *Mol Syst Biol* 2008, **4**:168.
- Wintermute EH, Silver PA: Emergent cooperation in microbial metabolism. *Mol Syst Biol* 2010, **6**:407.
- Klitgord N, Segrè D: Environments that induce synthetic microbial ecosystems. *PLoS Comput Biol* 2010, **6**:e1001002.
- Phalakornkule C, Lee S, Zhu T, Koepsel R, Ataai MM, Grossmann IE, Domach MM: A MILP-based flux alternative generation and NMR experimental design strategy for metabolic engineering. *Metab Eng* 2001, **3**:124–137.
- Fischer E, Sauer U: Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nat Genet* 2005, **37**:636–640.
- Lestas I, Vinnicombe G, Paulsson J: Fundamental limits on the suppression of molecular fluctuations. *Nature* 2010, **467**:174–178.
- Segrè D, Vitkup D, Church GM: Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A* 2002, **99**:15112–15117.
- Boyd SP, Vandenberghe L: *Convex optimization*, Illustrated, reprint. Cambridge, UK; New York: Cambridge University Press; 2004.
- Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BØ: Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 2004, **429**:92–96.
- Hoppe A, Hoffmann S, Holzhütter H-GG: Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst Biol* 2007, **1**:23.
- Noor E, Lewis NE, Milo R: A proof for loop-law constraints in stoichiometric metabolic networks. *BMC Syst Biol* 2012, **6**:140.
- Kuepfer L: Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res* 2005, **15**:1421–1430.
- Kim J, Reed JL: RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Genome Biol* 2012, **13**:R78.
- Segrè D, Zucker J, Katz J, Lin X, D'haeseleer P, Rindone WP, Kharchenko P, Nguyen DH, Wright MA, Church GM: From annotated genomes to metabolic flux models and kinetic parameter fitting. *OMICS* 2003, **7**:301–316.
- Nakahigashi K, Toya Y, Ishii N, Soga T, Hasegawa M, Watanabe H, Takai Y, Honma M, Mori H, Tomita M: Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. *Mol Syst Biol* 2009, **5**:306.
- Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H: Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the keio collection. *Mol Syst Biol* 2006, **2**:2006. 0008.
- Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A: Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science* 2007, **316**:593.
- Meng X-L, Rosenthal R, Rubin DB: Comparing correlated correlation coefficients. *Psychol Bull* 1992, **111**:172.
- Schuetz R, Kuepfer L, Sauer U: Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol Syst Biol* 2007, **3**:119.
- Zhao J, Baba T, Mori H, Shimizu K: Global metabolic response of *Escherichia coli* to *gnd* or *zwf* gene-knockout, based on 13 C-labeling experiments and the measurement of enzyme activities. *Appl Microbiol Biotechnol* 2004, **64**:91–98.
- Mustafa G, Migita CT, Ishikawa Y, Kobayashi K, Tagawa S, Yamada M: Menaquinone as well as ubiquinone as a bound quinone crucial for catalytic activity and intramolecular Electron Transfer in *Escherichia coli* membrane-bound glucose dehydrogenase. *J Biol Chem* 2008, **283**:28169–28175.
- Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U: Multidimensional optimality of microbial metabolism. *Science* 2012, **336**:601–604.
- Sariyar B, Perk S, Akman U, Hortacısu A: Monte Carlo sampling and principal component analysis of flux distributions yield topological and modular information on metabolic networks. *J Theor Biol* 2006, **242**:389.
- Wiback SJ, Famili I, Greenberg HJ, Palsson BØ: Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J Theor Biol* 2004, **228**:437–447.
- Andersen ED, Roos C, Terlaky T: On implementing a primal-dual interior-point method for conic quadratic optimization. *Math Program* 2003, **95**:249–277.
- Matheiss TH, Rubin DS: A survey and comparison of methods for finding all vertices of convex polyhedral sets. *Mathematics of operations research* 1980, **5**:167–185.
- Bordbar A, Lewis NE, Schellenberger J, Palsson BØ, Jamshidi N: Insight into human alveolar macrophage and *M. Tuberculosis* interactions via metabolic reconstructions. *Mol Syst Biol* 2010, **6**:422.
- Kaufman DE, Smith RL: Direction choice for accelerated convergence in hit-and-run sampling. *Oper Res* 1998, **46**:84–95.
- van Rijsewijk Bart RBH, Nanchen A, Nallet S, Kleijn RJ, Sauer U: Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Mol Syst Biol* 2011, **7**:1–12.
- Knorr AL, Jain R, Srivastava R: Bayesian-based selection of metabolic objective functions. *Bioinformatics* 2007, **23**:351–357.
- Gianchandani EP, Oberhardt MA, Burgard AP, Maranas CD, Papin JA: Predicting biological system objectives de novo from internal state measurements. *BMC Bioinformatics* 2008, **9**:43.
- Oh Y, Lee D, Lee S: Multiobjective flux balancing using the NISE method for metabolic network analysis. *Biotechnol Prog* 2009, **25**:999–1008.
- Zomorodi AR, Maranas CD: OptCom: a multi-level optimization Framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput Biol* 2012, **8**:e1002363.
- Bull JJ, Wang IN: Optimality models in the age of experimental evolution and genomics. *J Evol Biol* 2010, **23**:1820–1838.
- Pérez-Escudero A, Rivera-Alba M, de Polavieja GG: Structure of deviations from optimality in biological systems. *Proc Natl Acad Sci USA* 2009, **106**:20544–20549.
- Kitano H: Biological robustness. *Nat Rev Genet* 2004, **5**:826–837.
- Maltsev N, Glass EM, Ovchinnikova G, Gu Z: Molecular mechanisms involved in robustness of yeast central metabolism against null mutations. *J Biochem* 2005, **137**:177–187.
- Marland E, Prachumwat A, Maltsev N, Gu Z, Li W-HH: Higher gene duplicabilities for metabolic proteins than for nonmetabolic proteins in yeast and *E. coli*. *J Mol Evol* 2004, **59**:806–814.
- Kitami T, Nadeau JH: Biochemical networking contributes more to genetic buffering in human and mouse metabolic pathways than does gene duplication. *Nat Genet* 2002, **32**:191–194.
- Fell D: *Understanding the control of metabolism*. London/Miami: Portland Press Ltd; 1997.
- Springer M, Weissman JS, Kirschner MW: A general lack of compensation for gene dosage in yeast. *Mol Syst Biol* 2010, **6**:1–8.
- Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS: Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 2010, **329**:533–538.
- Vilar JMG, Guet CC, Leibler S: Modeling network dynamics: the lac operon, a case study. *J Cell Biol* 2003, **161**:471–476.
- Veening J-WW, Smits WK, Kuipers OP: Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* 2008, **62**:193–210.
- Thattai M, Van Oudenaarden A: Stochastic gene expression in fluctuating environments. *Genetics* 2004, **167**:523.



54. Süel GURM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB: **Tunability and noise dependence in differentiation dynamics.** *Science* 2007, **315**:1716–1719.
55. Lee HH, Molla MN, Cantor CR, Collins JJ: **Bacterial charity work leads to population-wide resistance.** *Nature* 2010, **467**:82–85.
56. Ptashne M: *A genetic switch : phage lambda revisited.* 3rd edition. Cold Spring Harbor, N.Y.: CSHL Press; 2004.
57. Spencer SL, Sorger PK: **Measuring and modeling apoptosis in single cells.** *Cell* 2011, **144**:926–939.
58. Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R: **Direct cell reprogramming is a stochastic process amenable to acceleration.** *Nature* 2009, **462**:595–601.
59. Tyagi S: **E. coli, what a noisy bug.** *Science* 2010, **329**:518.
60. Okumoto S: **Imaging approach for monitoring cellular metabolites and ions using genetically encoded biosensors.** *Curr Opin Biotechnol* 2010, **21**:45–54.
61. Balázsi G, van Oudenaarden A, Collins JJ: **Cellular decision making and biological noise: from microbes to mammals.** *Cell* 2011, **144**:910–925.
62. Feist AM, Palsson BØ: **The biomass objective function.** *Curr Opin Microbiol* 2010, **13**:344–349.
63. Gass SI: *Linear programming : methods and applications.* 5th edition. New York: Courier Dover Publications; 1985.
64. Kozlov MK, Tarasov SP, Khachiyan LG: **Polynomial solvability of convex quadratic programming.** *Soviet Mathematics Doklady* 1979, **20**:1108–1111.

doi:10.1186/1752-0509-7-98

**Cite this article as:** Wintermute *et al.*: An objective function exploiting suboptimal solutions in metabolic networks. *BMC Systems Biology* 2013 **7**:98.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

