INTERFACE

Author for correspondence:
M. Ángeles Serrano
e-mail: marian.serrano@ub.edu

THE ROYAL SOCIETY
PUBLISHING

Mapping high-growth phenotypes in the flux space of microbial metabolism

Oriol Güell[1], Francesco Alessandro Massucci[3], Francesc Font-Clos[4,5], Francesc Sagués[1] and M. Ángeles Serrano[2]

[1]Departament de Química Física, and [2]Departament de Física, Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain
[3]Departament d'Enginyeria Química, Universitat Rovira i Virgili, 43007 Tarragona, Spain
[4]Centre de Recerca Matemàtica, Edifici C, Campus Bellaterra, Bellaterra, 08193 Barcelona, Spain
[5]Departament de Matemàtiques, Universitat Autònoma de Barcelona, Edifici C, Campus Bellaterra, Bellaterra, 08193 Barcelona, Spain

FF-C, 0000-0003-4938-2452

Experimental and empirical observations on cell metabolism cannot be understood as a whole without their integration into a consistent systematic framework. However, the characterization of metabolic flux phenotypes is typically reduced to the study of a single optimal state, such as maximum biomass yield that is by far the most common assumption. Here, we confront optimal growth solutions to the whole set of feasible flux phenotypes (FFPs), which provides a benchmark to assess the likelihood of optimal and high-growth states and their agreement with experimental results. In addition, FFP maps are able to uncover metabolic behaviours, such as aerobic fermentation accompanying exponential growth on sugars at nutrient excess conditions, that are unreachable using standard models based on optimality principles. The information content of the full FFP space provides us with a map to explore and evaluate metabolic behaviour and capabilities, and so it opens new avenues for biotechnological and biomedical applications.

1. Introduction

Growing evidence suggests that metabolism is a dynamically regulated system that reorganizes under evolutionary pressure to safeguard survival [1,2]. This adaptability implies that metabolic phenotypes directly respond to environmental conditions. For instance, unicellular organisms can be stimulated to proliferate by controlling the abundance of nutrients available. In rich media, cells reproduce as quickly as possible by fermenting glucose, a process that produces high specific growth rates as well as large quantities of excess carbon in the form of ethanol and organic acids [3], a process known as the Crabtree effect [4,5]. To survive the scarcity of nutrients during starvation periods, glycolysis is hypothesized to switch to oxidative metabolism, which no longer maximizes the specific growth rate, but instead the ATP yield needed for cellular processes. Cells of multicellular organisms show similar metabolic phenotypes, relying primarily on oxidative phosphorylation when not stimulated to proliferate and changing to non-oxidative glycolytic metabolism during cell proliferation, even if this process—known in cancer cells as the Warburg effect [6]—is much less efficient at the level of energy yield.

These metabolic phenotypes are captured by computational approaches such as flux balance analysis [7] (FBA) that has been applied to high-quality genome-scale metabolic network reconstructions [8–12] to estimate the fluxes of biochemical reactions at steady state. Compliant with stoichiometric mass balance constraints and with imposed upper and lower bounds for nutrients, FBA determines the flux distribution that optimizes a biological objective such as specific growth rate, biomass yield, ATP yield or the rate of production of a biotechnologically important metabolite. This important tool has been used to predict the growth rate of organisms and to analyse their viability [13,14].

Minimization of metabolic adjustment [15], which identifies a single suboptimal point in flux space, has been proposed as an alternative option for perturbed metabolic networks not exposed to long-term evolutionary pressure. In any case, the identified solutions are frequently inconsistent with the biological reality, because no single objective function describes successfully the variability of flux states under all environmental conditions [16,17], and in fact, the highest accuracy of FBA predictions is achieved whenever the most relevant objective function is tailored to particular environmental conditions according to the empirical evidence for a very specific metabolic phenotype. For instance, FBA maximization of growth rate, by far one of the most common assumptions, requires either a rich medium or a manual limitation of the oxygen uptake to a physiological enzymatic limit to mimic the observed fermentation of glucose to formate, acetate or ethanol typical of proliferative metabolism, while in minimal medium, optimization of growth rate relies primarily on oxidative phosphorylation, which increases ATP production converting glucose to carbon dioxide, as in starvation metabolism.

Alongside optimal metabolic phenotypes, there is, however, a whole space of possible states that are non-reachable by invoking optimality principles that prevent non-optimal or typical biological states. Optimization of a biological function in the absence of *a priori* biological justification, similar to what happens under conditions for proliferative or starvation metabolism, may indeed weaken *in silico* predictions. Elementary flux modes [18,19]—non-decomposable steady-state pathways through a metabolic network such that any possible pathway can be described as a non-negative linear combination of them—provide a view on the flux space without the requirement of any optimality function. However, calculation of all elementary flux modes for an entire network is computationally very demanding owing to the combinatorial explosion of their number with increasing size of the network [20]. For instance, the core metabolism of *Escherichia coli* in reference [21] consists of around 271 million elementary flux modes [22]. To overcome this handicap, recent advances avoid the comprehensive enumeration of elementary flux modes, using instead a sample of the available elementary flux mode solution space [23]. Even assuming that one is able to enumerate all elementary flux modes, it is, however, impossible to assess the likelihood of observing a given linear combination of them in a typical phenotype. Further, elementary flux modes cannot capture changes associated with reaction fluxes being capped for whichever physiological reason (electronic supplementary material, figure S6). In addition, owing to functional redundancy, the expansion of possible metabolic pathways in elementary flux modes is not unique. Therefore, enumeration of the elementary flux modes is not as insightful as characterizing the whole phenotypic space, albeit requiring a comparable computational complexity.

Here, we introduce an alternative approach that estimates directly the feasible flux phenotype (FFP) space using a mathematically well-characterized sampling technique which enables the analysis of feasible flux states in terms of their likelihood. We use it to confront optimal growth rate solutions with the whole set of FFPs of *E. coli* core metabolism in minimal medium. The FFP space provides a reference map that helps us to assess the likelihood of optimal and high-growth states. We quantitatively and visually show that optimal growth flux phenotypes are eccentric with respect to the bulk of states,

represented by the FFP mean, which suggests that optimal phenotypes are uninformative about the more probable states, most of them low growth rate. We propose FFP space eccentricity of experimental data as a standard tool to calibrate the deviation of optimal phenotypes from experimental observations. Finally, the analysis of the entire high-biomass production region of the FFP space unveils metabolic behaviours observed experimentally but unreachable by models based on optimality principles, which forbid aerobic fermentation—a typical pathway utilization of proliferative metabolism—in minimal medium with unlimited oxygen uptake.

2. Material and methods

The FFP space, also termed the flux cone [24], of a metabolic model in a specific environment has been explored using different sampling techniques [25–27]. Here, we use the hit-and-run (HR) algorithm to explore the FFP space, tailoring it to enhance its sampling rate and to minimize its mixing time. We refer the interested reader to reference [28], where our implementation was first introduced, stating here only the key points and ideas.

2.1. Hit-and-run algorithm to sample the space of feasible metabolic flux solutions

We start by noting that all points in the FFP space must simultaneously satisfy mass balance conditions and uptake limits for internal and exchanged metabolites, respectively. The former requirement defines a set of homogeneous linear equalities, whose solution space is K, whereas the latter defines a set of linear inequalities, whose solutions lie in a convex compact set V. From a geometrical point of view, the FFP space is thus given by the intersection $S = K \cap V$. A key step of our approach consists of realizing that one can directly work in S by sampling V in terms of a basis spanning K. This allows retrieval of all FFPs that satisfy mass balance in the medium conditions under consideration, without rejection. Additionally, sampling in S enables a drastic dimensional reduction to be performed and considerably decreases the computation time. Indeed, assuming that there are N reactions, I internal metabolites and E exchanged metabolites ($N > I + E$), one has that $S \subset \mathbb{R}^{N-I}$, which is typically a space with greatly reduced dimensionality with respect to $V \subset \mathbb{R}^N$.

Once a basis for K is found, the main idea behind HR is fairly simple. Given a feasible solution $v_o \in S$, a new, different feasible solution $v_n \in S$ can be obtained as follows

(1) Choose a random direction u in $\mathbb{R}^I$,
(2) Draw a line ℓ through v_o along direction u:

$$\ell: \mathbf{v_o} + \lambda \mathbf{u}, \qquad \lambda \in \mathbb{R}$$

(3) Compute the two intersection points of ℓ with the boundary of S, parametrized by $\lambda = \lambda_-, \lambda_+$

$$\mathbf{v_-} = \mathbf{v_o} + (\lambda_-)\mathbf{u}$$
$$\mathbf{v_+} = \mathbf{v_o} + (\lambda_+)\mathbf{u}$$

(4) Choose a new point $\mathbf{v_n}$ from ℓ, uniformly at random between $\mathbf{v_-}$ and $\mathbf{v_+}$. In practice, this implies choosing a value λ_n in the range (λ_-, λ_+) uniformly at random, and then

$$\mathbf{v_n} \equiv \mathbf{v_o} + \lambda_n \mathbf{u}.$$

This procedure is repeated iteratively so that, given an initial condition, the algorithm can produce an arbitrary number of feasible solutions (see electronic supplementary material, figure S4 for an illustrative representation of the algorithm).

The initial condition, which must be a feasible metabolic flux state itself (i.e. it must belong to S), is obtained by other methods. We used and recommend MinOver, see [28,29], but any other technique is valid. In particular, in cases where small samples of the FFP space have already been obtained by other sampling techniques, such points can be used to feed the HR algorithm and produce a new, larger sample.

It was proved that HR converges towards the uniform sampling of S [30], and we took several measures to ensure that this was the case in our implementation (electronic supplementary material, figure S4). For each model, we initially created samples of size 1.3×10^9, giving rise to a final set of 10^6 feasible solutions, uniformly distributed along the whole FFP space.

2.2. Principal component analysis of the profile correlation matrix

Compared with phenotypic optimization or, e.g. elementary flux modes, FFP sampling has the advantage of allowing the computation of reaction pair correlations. These may be exploited to detect how global flux variability emerges in the system through principal component (PC) analysis [31,32] and to quantify, in turn, the closeness of optimal phenotypes to the bulk of the FFP. In what follows, we briefly describe the method, whereas an illustrative example is provided in electronic supplementary material, figure S5.

To perform such a study, we start by writing down the matrix C_{ij} of correlations between all reaction pairs i, j. In doing this, we measure how much the variability of reaction flux v_i affects the flux v_j (and vice versa). In mathematical terms, for each pair of reactions i, j, we have

$$C_{ij} = \frac{\langle v_i v_j \rangle - \langle v_i \rangle \langle v_j \rangle}{\sqrt{\left(\langle v_i^2 \rangle - \langle v_i \rangle^2\right)\left(\langle v_j^2 \rangle - \langle v_j \rangle^2\right)}}, \quad (2.1)$$

where $\langle \ldots \rangle$ denotes an average over the sampled set and the denominator of the fraction is simply the product of the standard deviations of v_i and v_j. We plot such matrix in figure 1e.

Matrix C is real and symmetric by definition and, thus, diagonalizable. This means that, for every eigenvector ρ_κ, one has $\mathbf{C}\rho_\kappa = \lambda_\kappa \rho_\kappa$. Note that matrix C describes paired flux fluctuations, in a reference frame centred on the mean flux vector. The eigenvectors ρ_κ of C express, in turn, the directions along which such fluctuations are taking place. In particular, the eigenvectors ρ_1, ρ_2 associated with the first two largest (in modulo) eigenvalues dictate the two directions in space where the sampled FFP displays the greatest variability (see electronic supplementary material, figure S5). This implies that sampled phenotypes lie closer to the plane spanned by ρ_1 and ρ_2 than the ones produced by any other linear combination of C eigenvectors. Projecting all sampled FFP onto this plane thus allows a drastic dimensional reduction to be performed while retaining much of the original variability and enables direct graphical insight into where phenotypes lie, where the bulk of the FFP is located and how the FBA solution compares with them. In such a plot, each phenotype $\jmath$ is described by two coordinates that may be parametrized via a radius $r_\jmath$ and an angle $\theta_\jmath$. Because the projection is normalized, it follows that $r_\jmath \leq 1$. Furthermore, the closer $r_\jmath$ is to one, the better the phenotype $\jmath$ is described by only looking at variability along ρ_1, ρ_2. As $r_\jmath$ is one at most, and because we have so many phenotypes clustered together, we chose to plot the PCA projection by using an effective radius $r_\jmath' = -\log(r_\jmath)$ in figure 1f. In this way, we could better discriminate among different phenotypes and obtain a 'closest to the origin, closest to the ρ_1, ρ_2-plane' set-up.

When compared with previous work focused on characterizing the PCs of the solution space to obtain a low-dimensional decomposition of the steady flux states of the system [33], our approach presents two main conceptual differences. First, the sampling method used here produces a uniform sample over the full set of feasible flux states without introducing any bias towards high-growth flux states. Second, we aim at a full description of all feasible flux states to conduct a statistical analysis of feasible phenotypes, which cannot be done by only retaining PCs. We use PC analysis to visualize the eccentricity of the FBA solution, but for all other purposes, we take into account the whole set of metabolic states.

3. Results

We study the full metabolic flux space of the *E. coli* core metabolic model [21,7], a condensed version of the genome-scale metabolic reconstruction *i*AF1260 [10] that contains 73 central metabolism reactions and 72 metabolites. This network is complemented by the biomass formation reaction and the ATP maintenance reaction. As in FBA, feasible flux states of a metabolic network are those that fulfil stoichiometric mass balance constraints together with imposed upper and lower bounds on the reaction fluxes. These constraints restrict the number of solutions to a compact convex set which contains all possible flux steady states in a particular environmental condition. In glucose minimal medium, the FFP space of *E. coli* core metabolism is determined by 70 potentially active reactions, including biomass formation and ATP maintenance reaction, and 68 metabolites. Note that we allow negative values for reversible reactions. We apply a fast and efficient hit-and-run algorithm [28] (see Material and methods) that explores the full solution space at random to produce a raw sample of 10^9 feasible states from which we extract a final uniform representative set of 10^6 feasible states.

Note that our approach is suitable for genome-scale network sizes beyond the reduced size of the *E. coli* core model. There is not any fundamental or technical bottleneck that prevents its application to complete metabolic descriptions at the cell level, because uniform samples can also be generated in genome-scale networks. We used the *E. coli* core metabolism owing to a matter of computational time and ease of visualization.

3.1. Optimal growth is eccentric with respect to the full feasible flux phenotype space

From the sampled set of *E. coli* core metabolic states in minimal medium of glucose bounded to $10 \, \text{mmol}/(\text{gDW} \cdot \text{h})$, we collected the metabolic flux profile of each individual reaction as the set of its feasible metabolic fluxes. From this profile, we computed the probability density function $f(v)$ which describes the likelihood for a reaction to take on a particular flux value. As an example, see figure 1a for the biomass function. We observe a variety of shapes (electronic supplementary material, figure S1), all of them low-variance, most displaying a maximum probability for a certain value of the flux inside the allowed range (note that none of these histograms can have more than one peak owing to the convexity of the steady-state flux space), and many being clearly asymmetric.

To characterize the dispersion of the possible fluxes for each reaction, we measured its coefficient of variation $\text{CV}(f(v))$ calculated as the ratio between the standard deviation of possible fluxes and their average (electronic supplementary material, table S1). For all but three reversible reactions
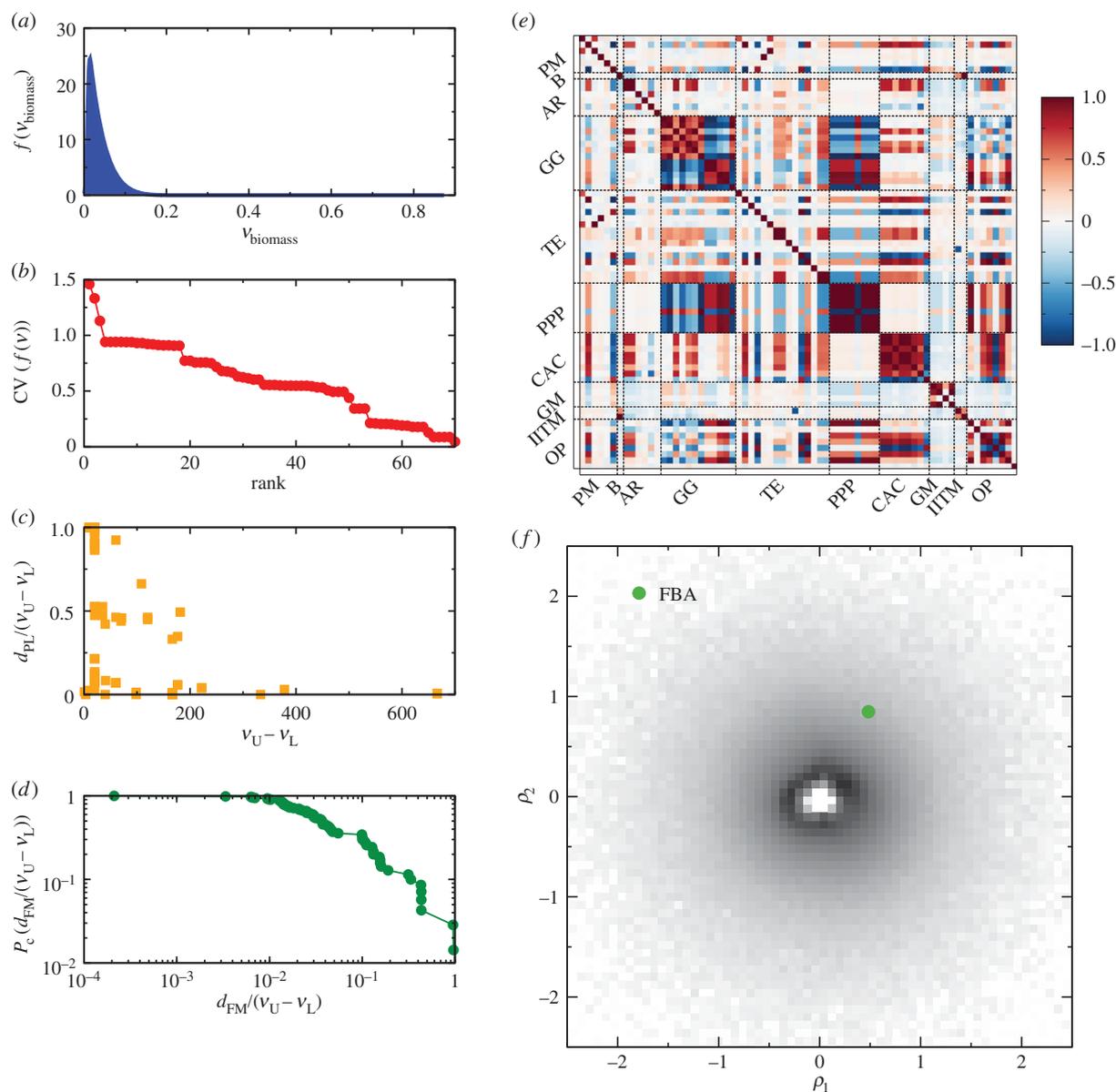
Figure 1. Analysis of reaction profiles and visualization of the FFP space. (*a*) Probability density function of metabolic flux values for the biomass function in *E. coli* core metabolism under glucose minimal conditions. (*b*) Coefficient of variation for all core reactions ranked by value. (*c*) Scatterplot of distances between the more probable flux in the FFP space and the lower flux bound rescaled by flux variability range for each reaction. (*d*) Complementary cumulative distribution function of distances between FBA maximal growth flux and FFP space mean flux rescaled by flux variability range for each reaction, in log–log scale. (*e*) Matrix of Pearson correlation coefficients measuring the degree of linear associations between feasible fluxes of reactions (acronyms of the pathways are shown in electronic supplementary material, table S2). (*f*) Projection of the FFP space onto the two principal component vectors of the correlation matrix in (*e*). All sampled flux phenotypes are normalized and projected along the first (ρ_1) and second (ρ_2) PCs. The plot is in polar coordinates, with the negative logarithm of the radius. The majority of points lie in a circle close to the origin (the darker area). The FBA solution (green circle) is, conversely, rather eccentric. (Online version in colour.)

(malate dehydrogenase, glucose-6-phosphate isomerase and glutamate dehydrogenase), the only reversible reactions having a low associated flux mean and thus a higher CV($f(v)$), this metric is below one and when ranked for all reactions, it steadily decreases to almost zero (figure 1*b*). Interestingly, we find that this coefficient is significantly anticorrelated with the essentiality of reactions, as observed experimentally [34] (point-biserial correlation coefficient −0.29 with *p*-value 0.01). This means that essential reactions tend to have a highly concentrated profile of feasible fluxes. Besides, and only for the glucose transferase reaction GLCpts, we find a zero probability of having a zero flux, which is not surprising as the lower bound given by FVA is strictly greater than zero indicating that this reaction is essential for *E. coli* core metabolism in glucose minimal medium. The asymmetry of each profile was characterized

by the distance between the more probable flux in the FFP space and the lower flux bound of the flux variability range rescaled by the flux variability range of the reaction (electronic supplementary material, table S1). In figure 1*c*, we show a scatterplot of values for all 68 core reactions. Strikingly, the rescaled distances cluster in three regions around 0, 0.5 and 1 forming groups of sizes 38, 15 and 17, respectively. This indicates that the most probable flux is close to either the lower or upper bound or, conversely, the probability distribution function tends to be quite symmetric. Moreover, we also observe an anticorrelation between the length of the flux range and the position of the most probable flux, so that the closer this is to its maximum value the shorter the allowed range of fluxes.

In order to assess the likelihood of FBA maximization of the biomass reaction (FBA-MBR; or equivalently of the

growth rate) solutions in relation to typical points within the whole FFP space (typical, in our mathematical/computational context, means statistically representative in relation to the whole set of flux states contained in the FFP space), we calculated the average flux value for each reaction, that we named the mean, and compared it with the FBA optimal biomass production flux. The complementary cumulative distribution function of the distances between these two characteristic fluxes rescaled by the flux variability range of reactions is shown in figure 1d (electronic supplementary material, table S1). We observe a broad distribution of values over several orders of magnitude with no mean value that is actually very close to the FBA maximal solution except for a few reactions, which typically work at maximum growth. At the other end of the spectrum, deviated reactions include, for instance, excretion of acetate and phosphate exchange. As a summary, we conclude that the mean and the FBA biomass optimum are rather distant, which suggests that FBA optimal states are uninformative about phenotypes in the bulk of states in the FFP space.

To visualize neatly the eccentricity of the FBA maximum growth state with respect to the bulk of metabolic flux solutions, we used PC analysis [31,32] in order to reduce the high dimensionality of the full flux solution space projecting it onto a two-dimensional plane from the most informative viewpoint (see Material and methods). We took reaction profiles in pairs to calculate the matrix of Pearson correlation coefficients measuring their degree of linear association (figure 1e; electronic supplementary material, table S3). Note that ordering of reactions by pathways allows clear visual feedback of intra- and interpathway correlations taking place in the core metabolic network, such that clusters of highly correlated reactions appear as bigger darker squares. The two axes of our projection correspond to the two first PCs of this profile correlation matrix ρ_1 and ρ_2, which account for most of the variability in profile correlations. Each sampled metabolic flux state was rescaled as a z-score centred around the mean and projected onto these axes, as shown in the scatterplot, figure 1f, in polar coordinates, where we applied a negative logarithmic transformation to the radial coordinate for ease of visualization. We see that the majority of phenotypes have a radius close to zero. Because points closer to the origin are better described by the two PCs, this implies that ρ_1 and ρ_2 capture the largest variability of the sampled FFP. Clearly, the FBA optimal growth solution is rather eccentric with respect to typical solutions, with an associated radius of 0.98 in this representation. In fact, 97% of states have a smaller radius than the optimal growth solution (see electronic supplementary material, figure S2).

3.2. The feasible flux phenotype space gives a benchmark to calibrate the deviation of optimal phenotypes from experimental observations

We focus on the relationship between primary carbon source uptakes and oxygen need to illustrate the potential of the FFP space as a benchmark to calibrate the deviation of in silico predicted optimal phenotypes from experimental observations. Sampled FFP states of the E. coli core model, in particular FFP mean values as a function of the upper bound uptake rate of the carbon source, are compared with reported experimental data for oxygen uptakes in minimal medium
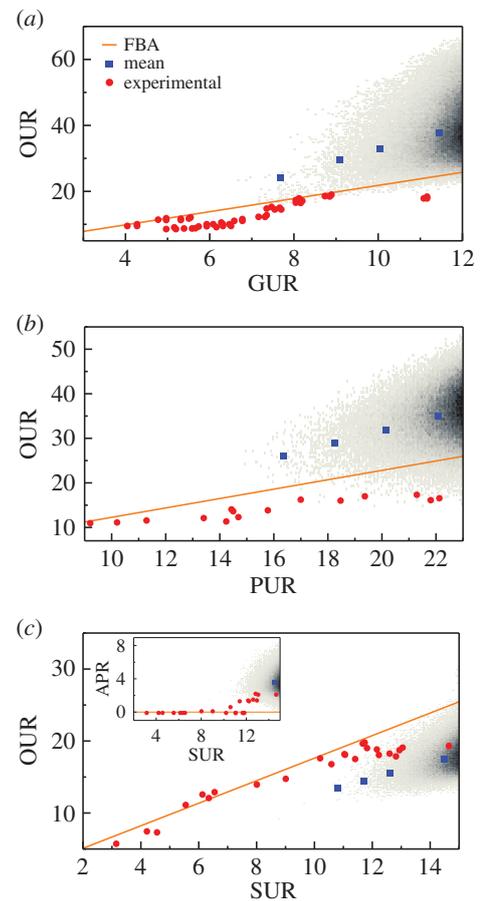
Figure 2. Comparison of predicted phenotypes and experimental data. Sampled points in the FFP space with maximum carbon source upper bound are plotted in shaded grey, darkness is proportional to the number of points. Experimental data points are red circles. The in silico-defined line of optimality, representing FBA optimal growth solutions as a function of the upper bound uptake rate of the carbon source, is shown in orange. Blue squares correspond to FFP mean values for different carbon source upper bound uptake rates. (a) Oxygen versus glucose uptake rates, experimental data from reference [1]. The FFP space is sampled with glucose bounded to 12 mmol/(gDW·h). (b) Oxygen versus pyruvate uptake rates, experimental data from reference [35]. The FFP space is sampled with pyruvate bounded to 23 mmol/(gDW·h). (c) Oxygen versus succinate uptake rates, experimental data from [36]. The FFP space is sampled with succinate bounded to 15 mmol/(gDW·h). Inset: acetate production rate versus. succinate uptake rate, experimental data from reference [36]. (Online version in colour.)

with glucose, pyruvate or succinate as a primary carbon source (figure 2). We also included in the figures the line of optimality representing FBA optimal growth solutions. We used glucose experimental data points from reference [1], experimental results for pyruvate reported in reference [35] and experimental results in reference [36] for the quantitative relationship between oxygen uptake rate and acetate production rate as a function of succinate uptake rate.

In all cases, FBA-MBR reproduces well experimental data points in the low carbon source uptake region [36], where E. coli is, indeed, optimizing biomass yield. However, the oxygen uptake rate saturates after some critical threshold of carbon source uptake rate, which depends on the carbon source reaching a plateau, which, among other possibilities, could be explained by the existence of a physiological enzymatic limit in oxygen uptake that lessens the

capacity of the respiratory system [37]. The plateau levels are 18.8 ± 0.7 for glucose [36], 16.8 ± 0.4 for pyruvate [35] and 19.49 ± 0.78 mmol/(gDW·h) for succinate [36]. In this region of high carbon source uptake, FBA-MBR predicts an oxygen uptake overestimated by around 25% with respect to the values reported from experiments. While this amount is, in principle, large, the FFP space gives a standard that helps to calibrate it.

We measured the eccentricity of experimental observations as their distance to the FFP mean. For glucose, this value is 19.4, which makes the distance of 5.3 between the FBA-MBR prediction and experimental data relatively low (figure 2a). The distance of 8.2 between the FBA-MBR prediction and experimental data is slightly worse for pyruvate (figure 2b) in that case the eccentricity of experimental observations is 18.4. The disagreement between optimality predictions and experimental data is much more significant in the case of succinate (figure 2c) for which the eccentricity of experimental observations is only of 4.3, whereas the distance between the FBA-MBR prediction and experimental data is 5.4, meaning that the FFP mean is, indeed, more adjusted to observations. The case of acetate production for this carbon source is even more conspicuous (figure 2c inset). While FBA-MBR is still reproducing well the experimental results of no acetate production in the low succinate uptake region, it cannot predict production of acetate at any succinate uptake rate due to the fact that FBA-MBR in minimal medium with unlimited oxygen does not capture the enzymatic oxygen limitation. The FBA-MBR solution diverts resources to the production of ATP entirely through the oxidative phosphorylation pathway. Thus, it fails to reproduce experimental observations of acetate production in the region of high succinate uptake rates [36,38–40]. In contrast, most metabolic states in the FFP space are consistent with acetate production, so that, in this case, the FFP mean turns out to be a good predictor of the experimentally observed metabolic behaviour.

In summary, while FBA-MBR predictions seem accurate for low carbon source uptake rate states in minimal medium as seen previously [36], the experimental points diverge from the FBA-MBR prediction state when increased values of carbon source uptake rates are considered. Note that, in general, it is not straightforward to quantify the significance of the divergence. Here, we propose to use the FFP space as a reference standard. According to this calibration, remarkably we find that FBA optimal growth predictions of oxygen needs versus glucose, pyruvate or succinate uptake are worse the more downstream the position of the carbon source into catalytic metabolism. Using the *E. coli* core metabolism, we have checked that the ratio of the maximum ATP production rate to the maximum oxygen uptake (both calculated by FBA optimization of ATP production rate) for the three carbon sources glucose, pyruvate and succinate are respectively 2.9, 2.6 and 2.4, so this ratio decreases as more downstream in the catalytic metabolism. These results are consistent with values reported in reference [37]. FBA privileges energy production by diverting fluxes to oxidative phosphorylation providing maximum energy for growth, so that FBA should work worse the less effective the oxidation of the carbon source is for ATP synthesis. This can be explained in terms of departures of energy from substrate catabolism to functions other than growth, such as basal maintenance, which become more relevant in relative terms when compared with the total energy

production when the energy-to-redox ratios of the carbon substrate are lower [41].

3.3. High-biomass production feasible flux phenotype region displays aerobic fermentation in minimal medium with unlimited oxygen uptake

We resampled the high-growth metabolic region of the *E. coli* core metabolism FFP space in glucose minimal medium with a glucose upper bound of 10 mmol/(gDW·h), as in §2.1. We defined this region by setting a minimal threshold for the biomass production of ≥ 0.4 mmol/(gDW·h) [42], and produced a sampled with a final size of 10^5 states. We note that phenotypes in this high-growth sample remain very close to the biomass yield threshold owing to the exponential decrease in the number of feasible flux states with increased biomass production, as in the biomass flux profile in figure 1a.

In this region, we identified pathway utilization typical of proliferative microbial metabolism, even when considering a minimal medium and unlimited oxygen uptake. This metabolic behaviour is consistent with experimental data [43,36,1], but it is unreachable by FBA models based on optimality principles (unless optimization is accompanied by auxiliary constraints not assumed in standard FBA implementations, such as the solvent capacity constraint [42], or by modelization beyond stoichiometric mass balance, introducing, for instance, thermodynamically feasible kinetics or enzyme synthesis [44,45]). We checked that the by-products cannot be explained by FBA-MBR in minimal medium with unlimited oxygen supply because, in this optimization framework, metabolic fluxes are basically forced to ATP production through oxidative phosphorylation with excretion of CO_2 as waste. However, increasing the oxygen limitation in FBA-MBR results in secretion of formate, acetate and ethanol—in that order—with corresponding shifts in metabolic behaviour [37].

According to the FFP space of *E. coli* core metabolism, we observe that the high-biomass production FFP subsample is characterized by the secretion of small organic acid molecules, even when the supply of oxygen is unlimited. This fact points to the simultaneous utilization of glycolysis and oxidative phosphorylation to produce biomass and energy and so to suboptimal states. This observation is supported by results from ^{13}C-metabolic flux analysis in *E. coli* [46], where repressed oxidative phosphorylation was proposed as responsible for the measured submaximal aerobic growth. Pathway utilization is illustrated in the schematic shown in figure 3a. Quantitative relationships between the production of small organic acids molecules and glucose and oxygen uptake rates are shown in the remaining panels of figure 3. Three-dimensional scatterplots for the production rates of formate, acetate and ethanol are shown in figure 3b,d,f, respectively, with projections into the three possible two-dimensional planes shown in figure 3c,e,g, respectively. Electronic supplementary material, figure S3 gives results for lactate. As the levels of glucose and oxygen uptakes are raised, metabolic phenotypes can achieve an increased production of formate, acetate and ethanol, even though the majority of feasible phenotypes remain at low organic acids production values. Owing to the high-growth requirement, oxygen uptake is always high but its variability increases with glucose uptake increase around a value of approximately 41.2 mmol/(gDW·h), which clusters the majority of
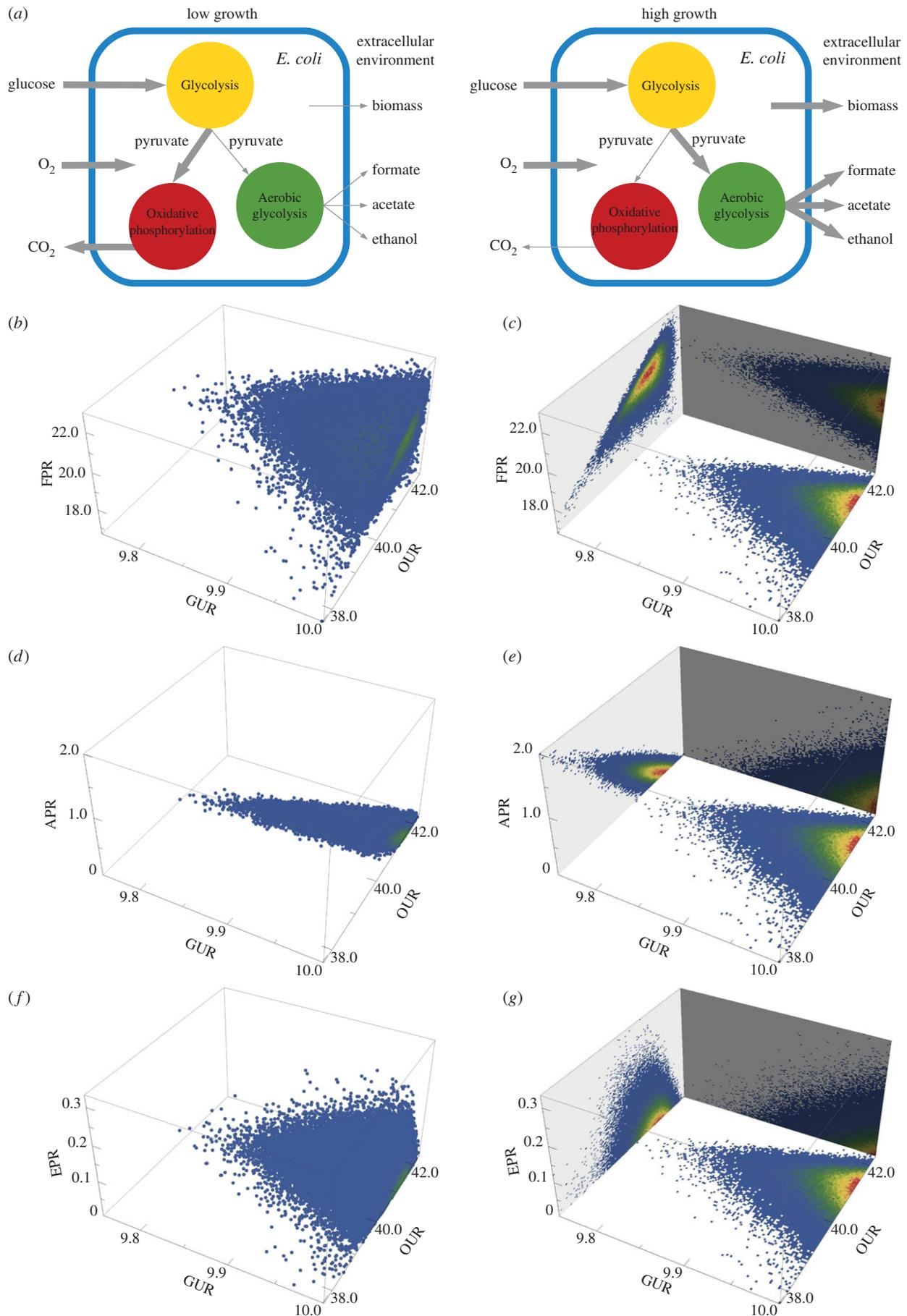
Figure 3. High growth phenotypes of *E. coli* core metabolism on glucose minimal medium. (*a*) Schematic of pathway utilization in high-growth versus low-growth conditions. (*b*) Three-dimensional scatterplot of formate production rate versus glucose and oxygen uptake rates. (*c*) Density projections of (*b*) on each of the possible two-dimensional planes, formate – glucose, formate – oxygen and glucose – oxygen. (*d*) Three-dimensional scatterplot of acetate production rate versus glucose and oxygen uptake rates. (*e*) Density projections of (*d*) on each of the possible two-dimensional planes, acetate – glucose, acetate – oxygen and acetate – oxygen. (*f*) Three-dimensional scatterplot of ethanol production rate versus glucose and oxygen uptake rates. (*g*) Density projections of (*f*) on each of the possible two-dimensional planes, ethanol – glucose, ethanol – oxygen and glucose – oxygen.

high-growth metabolic phenotypes. Interestingly, this oxygen uptake rate value marks a region in the FFP space with maximum potential production rates of formate, acetate and ethanol. Above and below that value, most states are concentrated in the range [39,42] mmol/(gDW·h).

Taken together, these results indicate that, contrarily to FBA predictions, a high level of glucose uptake combined with unlimited oxygen can maintain the requirements of proliferative metabolism for biomass formation through aerobic fermentation even in minimal medium. At the same time, additional oxygen uptake diverts glucose back towards more efficient ATP production through oxidative phosphorylation. Hence, oxygen has the potential of regulating the glucose metabolic switch in which glucose uptake rates larger than a critical threshold around 5.0 mmol/(gDW·h) [42] lead to a linearly increasing maximum production of organic acid by-products by gradual activation of aerobic fermentation and a slight decrease of oxidative phosphorylation. The reduction of glycolysis fluxes for oxygen-sufficient conditions was reported previously in the context of ^{13}C-metabolic flux analysis [46].

4. Discussion

The information content of the full FFP space of metabolic states in a certain environment provides us with an entire map to explore and evaluate metabolic behaviour and capabilities. While optimality goals need to be tailored to conditions and produce singular optimal solutions that may not be consistent with experimental observations, we have nowadays sufficient computational and methodological capacity to produce and analyse full FFP maps. The latter offer a reference framework to put into perspective the likelihood of particular phenotypic states that, as shown, enables elucidation of metabolic behaviours that are unreachable using models based on optimality principles. In fact, the location of metabolic flux distributions into precise optimal states has been challenged recently. Chen et al. [46] found submaximal growth in aerobic conditions when steady-state ^{13}C-metabolic flux analysis was applied to E. coli. It has been proposed that metabolic flux evolves under the trade-off between two forces, optimality under one given condition and minimal adjustment between conditions [17]. In this way, resilience to changing environments necessarily forces flux states to near-optimal, but suboptimal regions of feasible flux states in order to maintain adaptability.

In the FFP map of E. coli core metabolism in aerobic minimal medium, optimal growth states appear as eccentric and far from the bulk of more probable phenotypes represented by the FFP mean, which offers an ergodic perspective of the FFP space in which all states can be explored at random with equal probability. One of the uses of the method is precisely to evidence the effects of evolutionary pressure on organisms, which may actually result in eccentric flux states. On the other hand, the FFP space gives a standard to calibrate the deviation of optimal phenotypes from experimental observations. Oxygen consumption is a particularly interesting target for analysis because it has been identified as a trigger of metabolic shifts [37,47]. Interestingly, according to the FFP map as a reference standard, we found that in high-growth conditions FBA-MBR predictions of experimental observations for unlimited oxygen needs versus

glucose, pyruvate or succinate uptakes are worse the more downstream the uptake of the carbon source into the catalytic metabolic stream. This is consistent with the fact that the FBA-MBR solution diverts resources to the production of ATP entirely through the oxidative phosphorylation pathway, so that the greater the effective potential of the carbon source to recombine with oxygen to produce energy, the more convergent the in silico prediction and the observed states.

In order to correct FBA in high-growth conditions, some investigations restricted the solution space beyond mass balance and uptake bounds through additional thermodynamic, kinetic or physiological constraints, such as the solvent capacity constraint quantifying the maximum amount of macromolecules that can occupy the intracellular space [42]. Alternatively, the objective function implemented in FBA has been modified to nonlinear maximization of the ATP or biomass yield per flux unit [16]. Other models consider constraints beyond stoichiometric mass balance, for instance thermodynamically feasible kinetics or enzyme synthesis [44,45]. While these FBA modifications enhance some predictions, their effectiveness depends on the estimation of kinetic coefficients using empirical or experimental data. In contrast, the FFP map contains the set of all solutions determined solely on the basis of mass balance and upper and lower bounds for nutrients, and therefore, it includes solutions compliant with physiological constraints or with the limitations imposed by complex metabolic regulation [48]. In particular, the FFP space naturally displays all high-growth feasible states that show characteristic metabolic behaviours such as aerobic fermentation with unlimited oxygen uptake even in minimal medium without the need to determine additional constants. This aerobic fermentation, which is apparently inefficient in terms of energy yield when compared with oxidative phosphorylation but demonstrated as a favourable catabolic state for all rapidly proliferating cells with high glucose uptake capacity [42], turns out as a probable metabolic phenotype even in minimal medium. Results in this direction have been reported using steady-state ^{13}C-metabolic flux analysis, which has shown that E. coli grows suboptimally in glucose minimal media owing to limited oxidative phosphorylation [46].

Beyond theoretical implications, FFP maps of microbial organisms can be of particular interest as tools for biotechnological applications, for instance in the engineering of E. coli fermentative metabolism as a fundamental cellular capacity for valuable industrial biocatalysis [49]. In biomedicine, the investigation of FBA optimal phenotypes in the framework of the FFP map can help to contextualize disease phenotypes in comparison with normal states. For instance, FBA proved suitable for modelling complex diseases such as cancer as it assumes that cancer cells maximize growth searching for metabolic flux distributions that produce essential biomass precursors at high rates [50,51]. The analysis of the entire region of high-growth phenotypes will allow the attainment and study of a variety of suboptimal feasible flux states close to optimality but which cannot be reproduced by optimality principles, and so it opens new avenues for the understanding of general and fundamental mechanisms that characterize this disease across subtypes.

References

1. Ibarra RU, Edwards JS, Palsson BØ. 2002 *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420**, 186–189. (doi:10.1038/nature01149)

2. Blank LM, Kuepfer L, Sauer U. 2005 Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol.* **6**, R49. (doi:10.1186/gb-2005-6-6-r49)

3. Frick O, Whittmann C. 2005 Characterization of the metabolic shift between oxidative and fermentative growth in *Saccharomyces cerevisiae* by comparative ^{13}C flux analysis. *Microb. Cell Factor.* **4**, 30. (doi:10.1186/1475-2859-4-30)

4. De Deken RH. 1996 The Crabtree effect: a regulatory system in yeast. *J. Gen. Microbiol.* **44**, 149–156. (doi:10.1099/00221287-44-2-149)

5. Paczia N, Nilgen A, Lehmann T, Gätgens J, Wiechert W, Noack S. 2012 Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microb. Cell Factor* **11**, 122. (doi:10.1186/1475-2859-11-122)

6. Vander Heiden MG, Cantley LC, Thompson CB. 2009 Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033. (doi:10.1126/science.1160809)

7. Orth JD, Thiele I, Palsson BØ. 2010 What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248. (doi:10.1038/nbt.1614)

8. Overbeek R et al. 2005 The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**, 5691–5702. (doi:10.1093/nar/gki866)

9. Blattner FR et al. 1997 The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474. (doi:10.1126/science.277.5331.1453)

10. Feist A et al. 2007 A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3**, 121. (doi:10.1038/msb4100155)

11. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R. 2007 Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J. Biol. Chem.* **282**, 28 791–28 799. (doi:10.1074/jbc.M703759200)

12. Mo ML, Palsson BØ, Herrgård MJ. 2009 Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst. Biol.* **3**, 37. (doi:10.1186/1752-0509-3-37)

13. Almaas E, Oltvai ZN, Barabási AL. 2005 The activity reaction core and plasticity of metabolic networks. *PLoS Comput. Biol.* **1**, 0557–0563. (doi:10.1371/journal.pcbi.0010068)

14. Suthers PF, Zomorrodi A, Maranas CD. 2009 Genome-scale gene/reaction essentiality and synthetic lethality analysis. *Mol. Syst. Biol.* **5**, 301. (doi:10.1038/msb.2009.56)

15. Segrè D, Vitkup D, Church GM. 2002 Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl Acad. Sci. USA* **99**, 15 112–15 117. (doi:10.1073/pnas.232349399)

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

17. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U. 2012 Multidimensional optimality of microbial metabolism. *Science* **336**, 601–604. (doi:10.1126/science.1216882)

18. Schuster S, Hilgetag C. 1994 On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* **2**, 165–182. (doi:10.1142/S0218339094000131)

19. Schuster S, Fell DA, Dandekar T. 2000 A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* **18**, 326–332. (doi:10.1038/73786)

20. Acuña V, Marchetti-Spaccamela A, Sagot MF, Stougie L. 2010 A note on the complexity of finding and enumerating elementary modes. *Biosystems.* **99**, 210–214. (doi:10.1016/j.biosystems.2009.11.004)

21. Orth JD, Fleming RM, Palsson BØ. 2010 Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *EcoSal Plus.* (doi:10.1128/ecosalplus.10.2.1)

22. Jungreuthmayer C, Zanghellini J. 2012 Designing optimal cell factories: integer programing couples elementary mode analysis with regulation. *BMC Syst. Biol.* **6**, 103. (doi:10.1186/1752-0509-6-103)

23. Gebauer J, Schuster S, de Figueiredo LF, Kaleta C. 2012 Detecting and investigating substrate cycles in a genome-scale human metabolic network. *FEBS J.* **279**, 3192–3202. (doi:10.1111/j.1742-4658.2012.08700.x)

24. Schilling CH, Edwards JS, Letscher D, Palsson BØ. 2000 Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.* **71**, 286–306. (doi:10.1002/1097-0290(2000)71:4<286::AID-BIT1018>3.0.CO;2-R)

25. Price ND, Schellenberger J, Palsson BØ. 2004 Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophys. J.* **87**, 2172–2186. (doi:10.1529/biophysj.104.043000)

26. Wiback SJ, Famili I, Greenberg HJ, Palsson BØ. 2004 Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J. Theor. Biol.* **228**, 437–447. (doi:10.1016/j.jtbi.2004.02.006)

27. Almaas E, Kovács B, Vicsek T, Oltvai ZN, Barabási A-L. 2004 Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature* **427**, 839–843. (doi:10.1038/nature02289)

28. Massucci FA, Font-Clos F, De Martino A, Pérez Castillo I. 2013 A novel methodology to estimate metabolic flux distributions in constraint-based models. *Metabolites* **3**, 838–852. (doi:10.3390/metabo3030838)

29. Krauth W, Mezard M. 1987 Learning algorithms with optimal stability in neural networks. *J. Phys. A* **20**, L745–L752. (doi:10.1088/0305-4470/20/11/013)

30. Lovász L. 1999 Hit-and-run mixes fast. *Math. Program.* **86**, 443–461. (doi:10.1007/s101070050099)

31. Pearson KLIII. 1901 On lines and planes of closest fit to systems of points in space. *Philos. Mag. Ser. 6* **2**, 559–572. (doi:10.1080/14786440109462720)

32. Jolliffe I. 2002 *Principal component analysis*. New York, NY: Springer.

33. Barrett CL, Herrgard MJ, Palsson BØ. 2009 Decomposing complex reaction networks using random sampling, principal component analysis and basis rotation. *BMC Syst. Biol.* **3**, 30. (doi:10.1186/1752-0509-3-30)

34. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BO. 2011 A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* **7**, 535. (doi:10.1038/msb.2011.65)

35. Fong S, Marciniak JY, Palsson BØ. 2003 Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale *in silico* metabolic model. *J. Bacteriol.* **185**, 6400–6408. (doi:10.1128/JB.185.21.6400-6408.2003)

36. Edwards JS, Ibarra RU, Palsson BØ. 2001 *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* **19**, 125–130. (doi:10.1038/84379)

37. Varma A, Boesch BW, Palsson BØ. 1993 Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol.* **59**, 2465–2473.

38. Reiling HE, Laurila H, Fiechter A. 1985 Mass culture of *Escherichia coli*: medium development for low and high density cultivation of *Escherichia coli* B/r in minimal and complex media. *J. Biotechnol.* **2**, 191–206. (doi:10.1016/0168-1656(85)90038-0)

39. El-Mansi EM, Holms WH. 1989 Control of carbon flux to acetate excretion during growth of *Escherichia coli* in batch and continuous cultures. *J. Gen. Microbiol.* **135**, 2875–2883. (doi:10.1099/00221287-135-11-2875)

40. Wolfe AJ. 2005 The acetate switch. *Microbiol. Mol. Biol. Rev.* **69**, 12–50. (doi:10.1128/MMBR.69.1.12-50.2005)

41. Varman AM, He L, You L, Hollinshead W, Tang YJ. 2014 Elucidation of intrinsic biosynthesis yields using ^{13}C-based metabolism analysis. *Microb. Cell Factor* **13**, 42. (doi:10.1186/1475-2859-13-42)

42. Vázquez A *et al*. 2008 Impact of the solvent capacity constraint on *E. coli* metabolism. *BMC Syst. Biol.* **2**, 7. (doi:10.1186/1752-0509-2-7)

43. Varma A, Palsson BØ. 1994 Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* **60**, 3724–3731.

44. Molenaar D, van Berlo R, de Ridder D, Teusink B. 2009 Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol. Syst. Biol.* **5**, 323. (doi:10.1038/msb.2009.82)

45. Wortel MT, Peters H, Hulshof J, Teusink B, Bruggeman FJ. 2014 Metabolic states with maximal specific rate carry flux through an elementary flux mode. *FEBS J.* **281**, 1547–1555. (doi:10.1111/febs.12722)

46. Chen X, Alonso AP, Allen DK, Reed JL, Shachar-Hill Y. 2011 Synergy between ^{13}C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in *E. coli*. *Metab. Eng.* **13**, 38–48. (doi:10.1016/j.ymben.2010.11.004)

47. Losen M, Frolich B, Pohl M, Buchs J. 2004 Effect of oxygen limitation and medium composition on *Escherichia coli* fermentation in shake-flask cultures. *Biotechnol. Prog.* **20**, 1062–1068. (doi:10.1021/bp034282t)

48. Chubukov V, Gerosa L, Kochanowski K, Sauer U. 2014 Coordination of microbial metabolism. *Nat. Rev.* **12**, 327–340. (doi:10.1038/nrmicro3238)

49. Orencio-Trejo M, Fernández-Sandoval MT, Utrilla J, Huerta-Beristain G, Gosset G, Martinez A. 2010 Engineering the *Escherichia coli* fermentative metabolism. *Adv. Biochem. Eng. Biotechnol.* **121**, 71–107. (doi:10.1007/10_2009_61)

50. Price ND, Shmulevich I. 2007 Biochemical and statistical network models for systems biology. *Curr. Opin. Biotechnol.* **18**, 365–370. (doi:10.1016/j.copbio.2007.07.009)

51. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. 2011 Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **7**, 501. (doi:10.1038/msb.2011.35)