

# SCIENTIFIC REPORTS



OPEN

## A simple strategy guides the complex metabolic regulation in *Escherichia coli*

Giuseppe Facchetti<sup>1,2</sup>

Received: 13 January 2016

Accepted: 27 April 2016

Published: 10 June 2016

A way to decipher the complexity of the cellular metabolism is to study the effect of different external perturbations. Through an analysis over a sufficiently large set of gene knockouts and growing conditions, one aims to find a unifying principle that governs the metabolic regulation. For instance, it is known that the cessation of the microorganism proliferation after a gene deletion is only transient. However, we do not know the guiding principle that determines the partial or complete recovery of the growth rate, the corresponding redistribution of the metabolic fluxes and the possible different phenotypes. In spite of this large variety in the observed metabolic adjustments, we show that responses of *E. coli* to several different perturbations can always be derived from a sequence of greedy and myopic resiliencings. This simple mechanism provides a detailed explanation for the experimental dynamics both at cellular (proliferation rate) and molecular level (<sup>13</sup>C-determined fluxes), also in case of appearance of multiple phenotypes. As additional support, we identified an example of a simple network motif that is capable of implementing this myopic greediness in the regulation of the metabolism.

The understanding of the ability of microorganisms to respond to various external perturbations represents an important but open question within Biology. A considerable effort has recently been devoted to studying the effect of different genetic and environmental changes, both experimentally and from a theoretical point of view, e.g. by measuring genes expression, <sup>13</sup>C-based fluxes<sup>1–4</sup> and by reconstructed genome-scale metabolic networks<sup>5–7</sup>. Through this perturbation analysis we hope to discover a global picture of the regulatory machinery that control the cellular metabolism. For example, it has been found that microorganisms like *Escherichia coli* respond to a gene knockout by arresting their proliferation and by activating a large set of alternative pathways, the so called latent pathways<sup>8</sup>. Nevertheless, this “survival” response is only transient because, by re-routing the activation of the metabolic reactions, the microorganism is able to rescue its growing capability<sup>9,10</sup>. However, depending on the perturbation and on the nutrients availability, this proliferation recovery can be complete or only partial. An analysis of the published results considered in this work shows that, differently from what is usually assumed<sup>11</sup>, there is a significant fraction of cases (about 20%, see Supplementary Fig. S1 and ref. 9) in which the maximal value of the growth rate is not achieved. Therefore, although new experimental results from specific perturbations always provide useful and valuable information, it is also worth to consider the ensemble of the available knowledge and try to find a unifying explanation. Why are there perturbations for which the microorganism is unable to reach the maximal growth rate? Why after some knockouts does the microorganism adopt that redistribution of the active pathways even if it does not provide an optimal recovery of the metabolic function? Different definitions of optimality, like for instance Pareto optimality<sup>12</sup>, do not provide a description of the regulatory mechanism that guides each step of the response dynamics of the microorganism. On the other hand, in the computational methods available in the current literature, dynamics and regulation are introduced by imposing an external time-dependent process (for example the consumption of a limited amount of nutrients or the intracellular crowding<sup>13</sup>) or by setting extra constraints and parameters, which are obtained by additional experimental measurements<sup>11,14–16</sup>. Therefore, none of these methods provides a unifying and simple explanation for the complex metabolic response.

In order to find such a basic interpretation, our analysis consists of four steps: (I) from a general biological *rationale* that does not make use of any specific information from the available experiments, we formulated a

<sup>1</sup>Dept. Molecular and Statistical Physics, SISSA – International School for Advanced Studies, Trieste, Italy. <sup>2</sup>ICTP – International Centre of Theoretical Physics, Trieste, Italy. Correspondence and requests for materials should be addressed to G.F. (email: giuseppe.facchetti@jic.ac.uk)

simple hypothesis about a possible unifying principle; (II) we validated the proposed hypothesis by a stoichiometric reaction model: without any tuning of the parameter to fit the experimental data, we tested whether experimental results can be reproduced by this hypothesis, both as growth rate recovery and as fluxes redistribution; (III) once it has passed this validation, the criterion is used to predict and to derive an explanation for some interesting cases of metabolic responses; (IV) finally, we showed how a possible motif in the regulatory network can implement the proposed principle at the molecular level.

## Results

**From general biological considerations to the greedy hypothesis.** Our starting point is the following *rationale*: as already mentioned, genetic and environmental perturbations cause a transient activation of the non-essential metabolic pathways that are progressively silenced as the microorganism adapts to the new condition<sup>17</sup>. Indeed, it is known that a resiliencing after a knockout can lead to the recovery of some cellular functions such as the growth rate<sup>10</sup>. In a microorganism, the resiliencing of a reaction is achieved through different feedback/feedforward mechanisms such as, for instance, down-regulation of gene expression, the allosteric effect of a metabolite, deactivation of the enzyme by phosphorylation or dephosphorylation. Clearly, all these mechanisms rely only on the information that is available to the cell at the time of the resiliencing (a myopic view). Indeed, it is reasonable to assume that the cell does not know the final point of the entire adaptation, neither as growth rate nor as activated/silenced pathways. Based on this simple and plausible reasoning, we hypothesized the following greedy heuristic: among all the activated reactions, the regulatory machinery of the cell had evolved in order to preferentially select the resiliencing that triggers a high (greedy) instantaneous (myopic) proliferation advantage.

**Validation of the greedy hypothesis.** To validate this hypothesis, we tested how a recursive greedy resiliencing adjusts the central metabolism of *E. coli* after a wide set of different perturbations (see Supplementary Fig. S2 and Table S1). Starting from a Flux Balance Analysis (FBA) method recently developed<sup>18</sup>, we built a procedure that recursively and stochastically adopts a greedy choice among all possible resiliencings: *the higher the induced recovery of the growth rate, the higher the probability for the cell to adopt that resiliencing*. In particular, the following equation has been used for the estimation of probability of resiliencing reaction  $k$ :

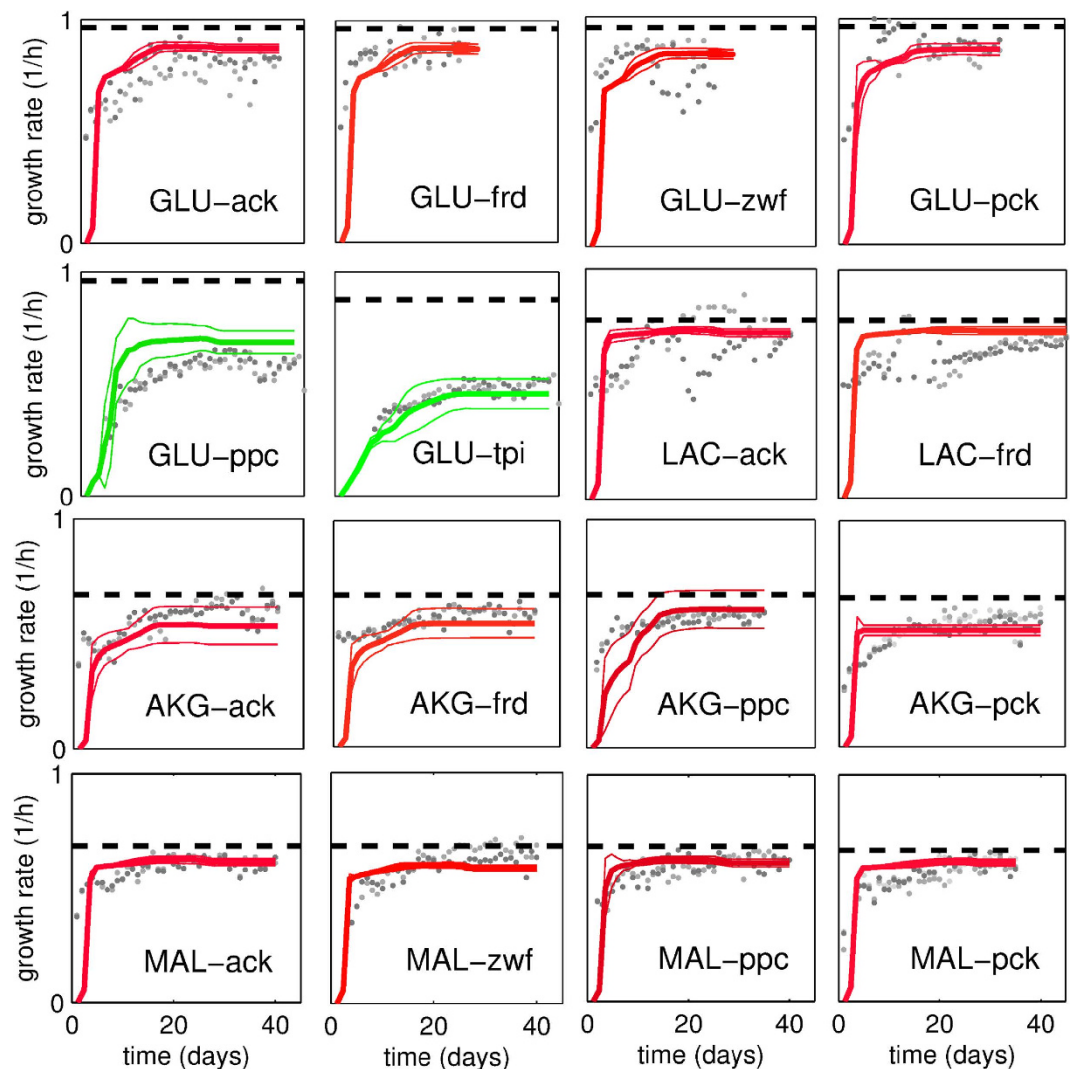
$$\mathbb{P}[k] = \frac{e^{g_k \beta}}{\sum_{j \in \mathcal{R}} e^{g_j \beta}}, \quad (1)$$

where  $g_j$  is the short-term recovery of the growth rate induced by resiliencing reaction  $j$  and obtained by MOMA<sup>6</sup>,  $\beta$  is the unique model parameter that describes the amount of randomness in the simulation and  $\mathcal{R}$  is the set of all active reactions, i.e. the set of all possible resiliencings (see Equations S6 and S7 in the Supplementary text for more details). From the recursive use of this resiliencing-MOMA procedure, we obtained a Markov Chain whose evolution generates a sequence of metabolic adjustments (Supplementary Fig. S3). We called the method GRAM: Greedy Resiliencing in the Adjustment of Metabolism. It is worth noting that the method does not use any additional constraints derived from transcriptional and regulatory information, and it does not introduce any artificial bias on the choice of the resiliencings. The entire dynamics are merely determined by the greedy criterion and by the given stoichiometry of the metabolic network. Moreover, because of the absence of any fitting procedure for the parameter  $\beta$  of the model (see Supplementary Fig. S4 for the choice of  $\beta = 200$  h) and because of the complete independence of the construction of the method from the specific experimental results it is asked to describe, we believe that the comparison between simulations and experimental data represents a stringent test on the validity of our hypothesis.

We considered 17 different combinations of single carbon sources media with gene knockouts on *E. coli*, namely glucose,  $\alpha$ -ketoglutarate, lactate, malate and succinate as carbon source and acetate kinase (*ack*), fumarate reductase (*frd*), glucose-6-phosphate dehydrogenase (*zwf*), phosphoenolpyruvate carboxylase (*ppc*), phosphoenolpyruvate carboxykinase (*pck*) and triose-phosphate isomerase (*tpi*) as deleted genes (see Supplementary Table S1 for the complete list). For each condition we simulated how the proposed greedy hypothesis would describe the metabolic regulation. The plots in Fig. 1 report the computed recoveries of the growth rate and the comparison with experimental data from ref. 9. For all substrates and knockouts, the entire experimental dynamics of the recovery of the growth rate follows very closely the trajectory provided by our greedy hypothesis. Of particular interest is the case of glucose as the carbon source: our criterion explains the fact that both  $\Delta tpi$  and  $\Delta ppc$  show a slower and only partial recovery of the growth rate, i.e. they do not reach the maximal value predicted through the classical biomass optimization (green colour). These results suggested then a general validity of the greedy hypothesis regardless of the knockout and growing condition.

It is worth noting that also in the case of high recovery, simulated growth rate does not reach exactly the maximal value suggested by classical FBA. Part of this small discrepancy might be due to the “calibration” of the nutrients availability (see Supplementary Fig. S2). However, we must also consider that, like for any optimization algorithm that differs from Linear Programming, there is no proof and guarantee that the proposed method should reach precisely the optimal solution given by the growth rate maximization. Indeed, consistent with our predictions, the experimental data also show that a truly complete recovery is not reached every time (see also Fig. 2b).

To further validate the greedy principle, we verified that the mathematical expression (1) used for greediness quantification was not stringent: as shown in Fig. 2a, the results indeed did not change when the  $m$ -th moment expression

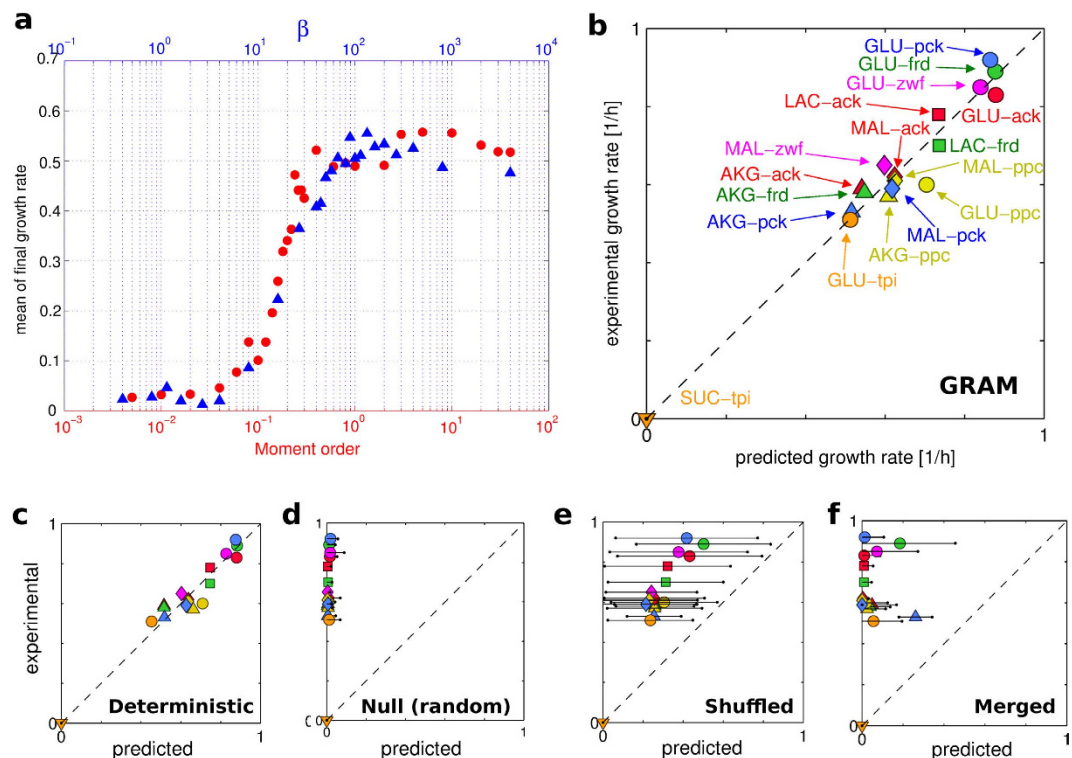


**Figure 1. Dynamics of the proliferation recovery of *E. coli* for different experimental conditions.** Each panel refers to the indicated pair carbon source-knockout and reports the results obtained with our criterion and the experimental data from ref. 9. *Legend.* AKG:  $\alpha$ -ketoglutarate, GLU: glucose, LAC: lactate, MAL: malate, SUC: succinate. Curves represent the results from GRAM (thick and thin lines: average and average  $\pm$  standard deviation). Red: complete or almost complete recovery ( $>90\%$ ); green: partial recovery ( $<90\%$ ). Gray dots: experimental values. Dashed black line: maximal growth rate (FBA biomass optimization). The SUC-*tpi* case has not been reported because it is lethal, i.e. a constant zero growth rate for both experiments and model.

$$\mathbb{P}[k] = \frac{(g_k)^m}{\sum_{j \in \mathcal{R}} (g_j)^m} \quad (2)$$

was used instead of the Boltzmann equation (1). Furthermore, the comparison with the FBA heuristics that are normally used showed our approach performed better, see Supplementary Fig. S5. Finally, we looked at how a deviation from greediness affects the quality of the results. In particular we tested a very greedy choice (purely deterministic, i.e. the cell always adopts the most convenient resiliencing), the hypothesis of a random choice and the effect of shuffling or merging the sequence of the greedy resiliencing we have identified. We ran each of these modified procedures for all the 17 experimental conditions. As shown in Fig. 2b–f, except for the pure deterministic version which still follows the greedy criterion, for all other deviations the agreement with the experimental data was significantly reduced or even lost. Therefore, also these tests supported our hypothesis.

So far, we found that this greedy regulation of the metabolism appears to be a good principle for explaining the macroscopic evidence of the growth rate recovery. This led us to the question about the re-routing of the metabolic pathways. Can the redistribution of the reaction fluxes be explained by the greedy hypothesis? For this purpose we looked into the details of the dynamics by identifying which reactions have been resilienced/activated along the adjustment and comparing this result with the  $^{13}\text{C}$ -determined fluxes at the intermediate steps and at the final steps of the adjustment in glucose for five different knockouts, namely *zwf*, *ppc* and *tpi*, plus phosphate



**Figure 2. Tests on the importance of greediness.** Comparison of the final growth rate obtained from some variation of greediness. (a) *GLU- $\Delta$ *tpi**: results by Boltzmann equation (1) (blue  $\blacktriangle$ , top *x* axis) and by moment equation (2) (red  $\bullet$ , bottom *x* axis) at different value of  $\beta$  and *m*, respectively. (b) Results from GRAM; RMSD of the mean = 0.0616. (c,d) Results from deterministic and random (null) models; RMSD of the mean = 0.0783 and 1.94 respectively. (e,f) Results from shuffled and merged sequence of the resiliencings obtained from the standard GRAM procedure; RMSD of the mean = 0.752 and 1.57, respectively). Growth rates in  $\text{h}^{-1}$ . Black horizontal lines: error bar (standard deviation). Color code for knockouts: red = *ack*, green = *frd*, violet = *zwf*, yellow = *ppc*, blue = *pck*, orange = *tpi*. Legend for carbon sources:  $\alpha$ -ketoglutarate (AKG,  $\Delta$ ), glucose (GLU,  $\circ$ ), lactate (LAC,  $\square$ ), malate (MAL,  $\diamond$ ) and succinate (SUC,  $\nabla$ ).

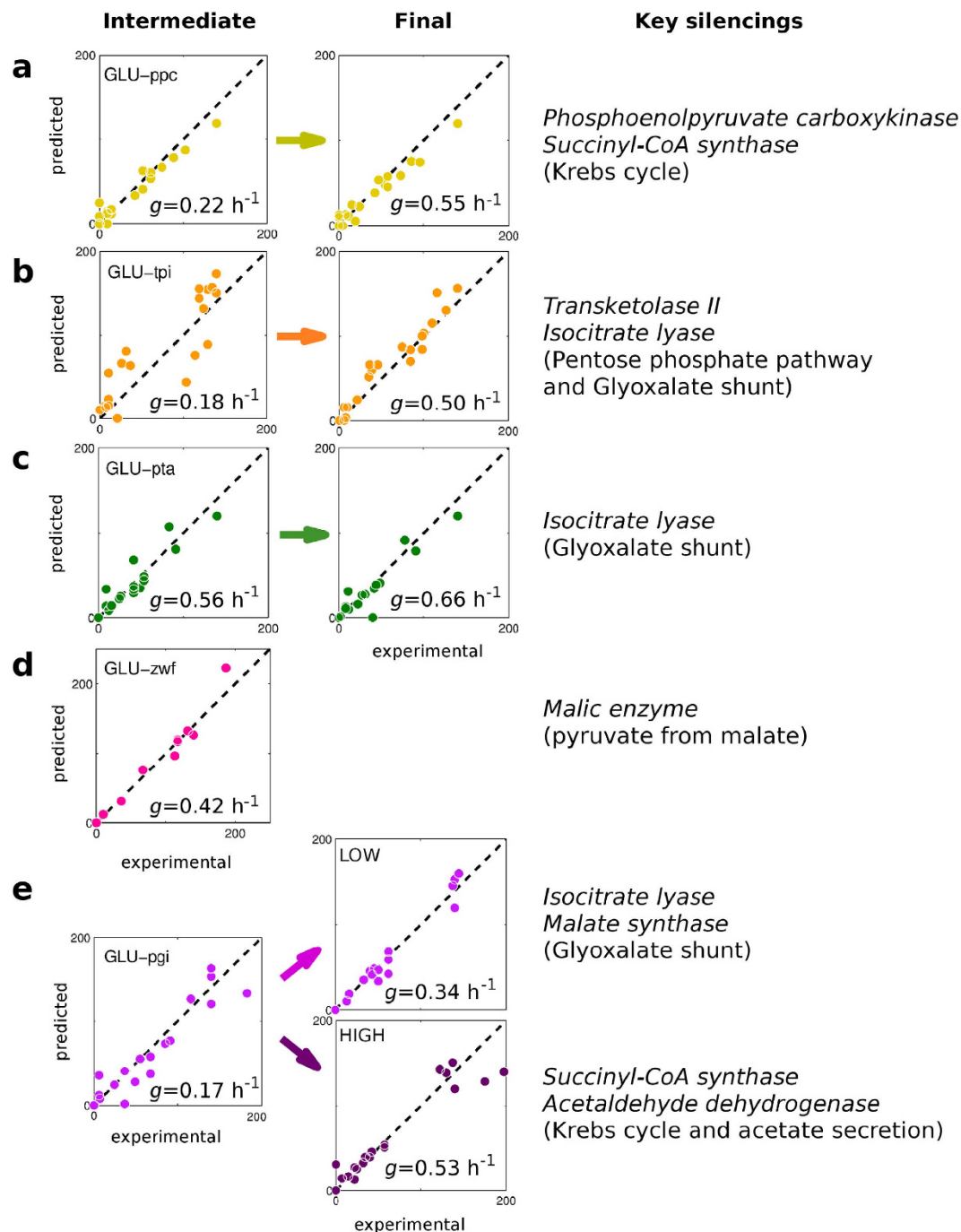
transacetylase *pta* and phosphoglucose isomerase *pgi*<sup>19,20</sup>. The results are displayed in Fig. 3 for intermediate and final points, together with the key resiliencings: all these predicted resiliencings were consistent with available <sup>13</sup>C-labelling experimental data. Therefore, for each knockout and regardless to which intermediate state along the trajectory has been used for the validation (validation points have a percentage of recovery that varies from 36% to 84%), the observed flux re-routing can be well described by our hypothesis. A comparison with other FBA methods is reported in Supplementary Table S2 and confirms the better performance of GRAM.

As indicated in Fig. 3e a significant example is the *pgi* knockout. It has been shown that this mutant is characterized by a reduction of the growth rate to approximately  $0.3 \text{ h}^{-1}$  and by a re-routing of the fluxes toward the pentose phosphate pathway<sup>21,22</sup>. However, in the experiments reported in ref. 19 a second phenotype with high growth rate ( $0.53 \text{ h}^{-1}$ ) has been identified. Measurements by <sup>13</sup>C-labelling indicated that the difference between these two phenotypes is due to the use (and non-use, respectively) of glyoxalate shunt and of acetate secretion. Surprisingly, these two possible growth rates were obtained also in our trajectories:  $0.31$  and  $0.60 \text{ h}^{-1}$  with about 70% and 20% probability, respectively. As indicated by Fig. 3e, computed fluxes are consistent with <sup>13</sup>C-measurements for both phenotypes (see Supplementary Fig. S7 for a sketch of the activated and silenced pathways in the two phenotypes). This agreement in the flux redistribution concerned also the adjustment of redox balance of the cell, i.e. the production or consumption of the NAD/NADH (see “Mutations” in Supporting Results). Therefore, even in the complex case of appearance of multiple phenotypes, the proposed principle is still able to explain the observed adjustments.

After having validated the hypothesis in terms of cellular growth rate and pathways re-routing, we looked at the importance of the temporal order in the regulation of the metabolism and how this order is captured by the greedy criterion. As reported above in Fig. 2e, an indirect proof was given by the loss of the prediction power when shuffling the sequence of resiliencing: by altering the order given by greediness, the cell becomes unable to correctly rescue its metabolic functions. Indeed, although some reactions can be stopped in a large time window, there are crucial reactions for which the temporal order must be preserved (see Supplementary Fig. S8). The two examples about the resiliencing of transketolase in *GLU-tpi* and about the resiliencing of membrane-bound NADPH transhydrogenase in *GLU-pgi* are described in the next paragraph.

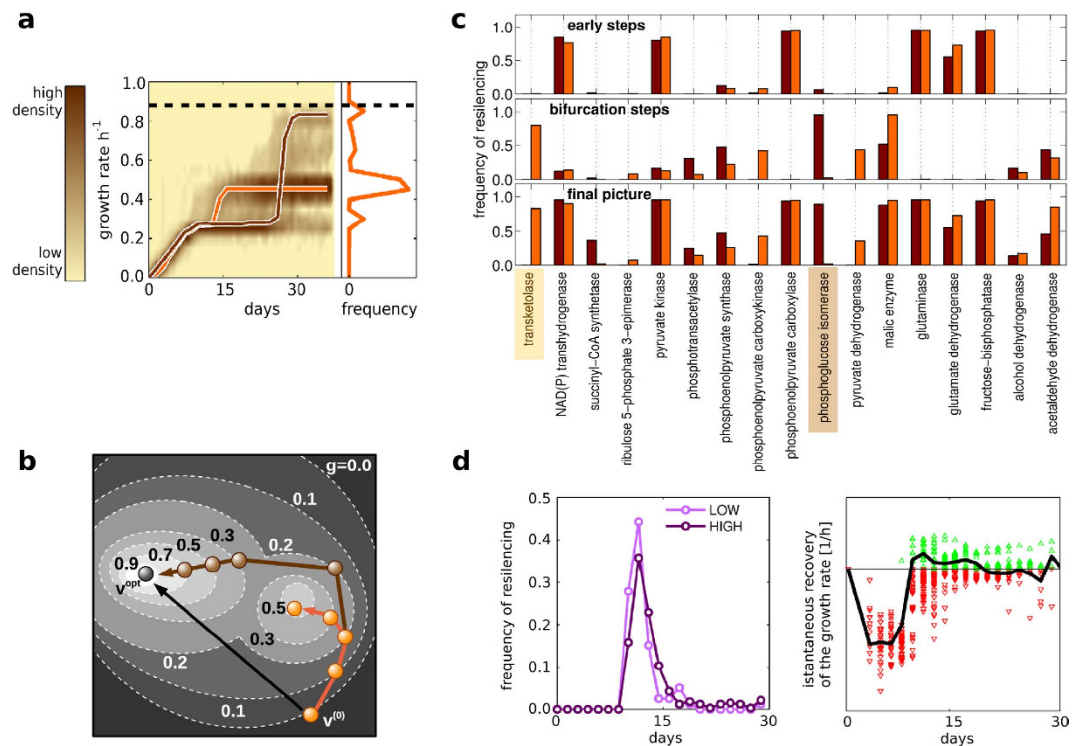
**Explanations from the greedy hypothesis.** With the detailed description provided by GRAM about the flux redistribution, we have now the possibility to decipher the mechanism that controls the metabolism.





**Figure 3. Comparison with  $^{13}\text{C}$ -labelling metabolic fluxes in glucose (from refs 19,20).** Experimental fluxes are compared with the flux calculated with the greedy hypothesis for intermediate and final steps of the adaptation for different knock-outs. The intermediate states along the trajectory have been chosen based on the growth rate of the corresponding  $^{13}\text{C}$ -labelling experiments used for the validation. Value of this growth rate is reported in each panel. (a) *GLU-ppc* (40% of recovery with respect to the final growth rate); (b) *GLU-tpi* (36% of recovery with respect to the final growth rate); (c) *GLU-ptg* (84% of recovery with respect to the final growth rate); (d) *GLU-zwf*, only experimental data for intermediate steps is available (48% of recovery with respect to the simulated end-point); (e) *GLU-pgi* (50% of recovery with respect to the final growth rate). In the latter case, the end-points for two different phenotypes are reported (denoted HIGH and LOW). For statistics and comparison with other methods see Supplementary Table S2. The last column reports the key silenced reactions (and pathways) predicted by GRAM and confirmed by the experimental measurements.

Our aim, indeed, is not to reproduce data that are already known from previous experiments, rather we want to provide a possible explanation for why such results have been observed. With this purpose, we present two significant examples in Fig. 4.



**Figure 4. Greediness provides explanation for sub-optimality and timing of the adjustments.** (a) *GLU-tpi* condition: density plot of the simulated time-recovery of the growth rate; the histogram at the right reports the frequency of the growth rates at the end-point. Two representative trajectories are shown explicitly: orange line refers to a trajectory with low final growth rate (0.48 h<sup>-1</sup>), brown line with high final growth rate (0.82 h<sup>-1</sup>). Same colours used in (b,c). (b) Interpretation of the dynamics of panel (a) as trajectories of a particle in an energy landscape where energy  $H = -g$ . (c) Frequency of resiliencing of critical reactions in the two *GLU-tpi* trajectories at the early, bifurcation and final steps (normalization with respect to the maximal value). Although at the early steps the two trajectories have similar flux redistributions, a crucial intermediate resiliencing (transketolase II is the greedy option, phosphoglucose isomerase is the non-greedy one) leads to different end-points. (d) *Left*: in *GLU-pgi* simulation, resiliencing of membrane NADPH transhydrogenase always occurs after 7 days (for LOW and HIGH phenotype, see Fig. 3e). *Right*: dots represent the instantaneous effect on the proliferation recovery depending on the resiliencing time of membrane NADPH transhydrogenase (green  $\Delta$ : positive recovery; red  $\nabla$ : negative recovery). Each dot refers to a time point of a single trajectory. Black curve represents the average.

As seen before,  $\Delta tpi$  is a very interesting perturbation after which the microorganism reaches only a partial recovery of the growth rate up to  $0.50 \pm 0.02 \text{ h}^{-1}$  (ref. 20), which is only half of the optimal value  $0.87 \text{ h}^{-1}$  obtained from the biomass maximization in FBA. Both GRAM results and <sup>13</sup>C-data indicate that the microorganism stops the pentose phosphate pathway and uses glycolysis as the main catabolic route<sup>19</sup>. Therefore, we would like to find the crucial step of the adjustment that explains this sub-optimal strategy and by which entity it fits into the greedy hypothesis. The trajectories of our model show a dominant adaptation strategy with a final growth rate equal to  $0.48 \text{ h}^{-1}$  (see orange line in Fig. 4a). Nevertheless, with a much lower probability (only 15%), the simulations show also a second set of trajectories which instead reach a growth rate of about  $0.82 \text{ h}^{-1}$  (brown line in Fig. 4a). The bifurcation between these two dynamics occurs after about 2 weeks. By looking at the resiliencing adopted at these intermediate steps, we found that transketolase II on the pentose phosphate pathway is always resilienced in the most frequent trajectories that reach low proliferation (orange bars in Fig. 4b), whereas it is never resilienced in the rare trajectories that end with high proliferation and for which phosphoglucose isomerase is instead stopped (brown bars in Fig. 4b). Moreover, we can also see that the resiliencing of transketolase II provides a short-term recovery of the biomass production up to  $0.45 \text{ h}^{-1}$  (see jump of orange curve after 15 days in Fig. 4a) which is higher than the alternative resiliencing of phosphoglucose isomerase in the glycolysis pathway (up to  $0.25 \text{ h}^{-1}$ , see brown curve in Fig. 4a after 15 days). Therefore, given these two possible resiliencings, our criterion suggests that the microorganism would preferably choose the resiliencing which assures a higher short-term recovery, i.e. it stops transketolase II in the pentose phosphate pathway and keeps glycolysis as the active metabolic route. Unfortunately, since the *tpi* knockout reduces the global long-term efficacy of glycolysis (unknown to the myopic cell at this bifurcation point), the final outcome is a phenotype with low proliferation rate. Figure 4c shows an interpretation in terms of an energy landscape, where the moving particle is attracted and trapped in a local minimum.

The fact that the highly proliferating phenotype has not been observed can then be interpreted as a confirmation of the existence of the greedy strategy (i.e. the low probability for the resiliencing of phosphoglucose

isomerase): if the cells do not follow the greedy criterion, it would have been easier to find this high proliferation phenotype. Moreover, as seen in the trajectories in Fig. 4a, there is a long transient dormant state in the non-greedy dynamics (brown curve) that might further reduce the experimental detectability of the high proliferation phenotype: indeed, after the silencing of phosphoglucose isomerase, the growth rate remains at a low value ( $0.25 \text{ h}^{-1}$ ) for about 10 days making these cells unable to compete for the common nutrients against the greedy cells that are transiently proliferating faster ( $0.45 \text{ h}^{-1}$ , orange curve).

Concerning the resiliencing time, Fig. 4b shows that in this mutant some reactions can be resilienced almost at any time of the dynamics (see for example pyruvate kinase, glutaminase and fructose-biphosphatase) whereas reactions which are crucial for the bifurcation (transketolase II and phosphoglucose isomerase) must be stopped only at the intermediate stage of the dynamics.

Finally, the use of GRAM provided also an interesting explanation for the apparent inconsistency regarding the resiliencing of the membrane-bound NADPH transhydrogenase in the *GLU-pgi* condition. Indeed, although a mutation that stops this reaction has been found in the adaptation of *GLU-pgi*<sup>23</sup>, it is not known why the insertion of the same mutation (i.e. at time zero) showed a clear disadvantage for the microorganism: the final growth rate was only  $0.1 \text{ h}^{-1}$ , instead of  $0.3$  or  $0.6 \text{ h}^{-1}$  (ref. 23). The exact time of this adaptation step is not known. In our trajectories it was instead possible to see that membrane-bound NADPH transhydrogenase was silenced approximately after 11 days and never before 7 days (Fig. 4d). Moreover, the same results showed that the effect of this resiliencing was always negative during the first week (red points in Fig. 4d). Therefore, the greedy interpretation suggests that, only once other adjustments have already been adopted, does the regulation of the NADPH metabolism become convenient. This is the reason why a forced stop of this reaction at time zero was not beneficial for the microorganism.

**An example of a greedy regulatory motif.** As mentioned, it is known that a cell can resiliencing a reaction through different feedback/feedforward mechanisms that down-regulate gene expression, have an allosteric effect on the enzymatic activity, or phosphorylate/dephosphorylate a protein. Therefore, the following question might arise: how can these mechanisms implement the proposed greedy strategy in a cell? To answer this question, we showed how a simple negative feedback motif, which is common in the regulatory networks, is capable of performing a greedy choice. Clearly, the set of feedbacks we present here is just an example and other alternative motifs with greedy behaviour may exist.

We used this motif in a minimal model. Because of its simplicity, local information coincides with global information, so that a greedy strategy always leads to the optimal point. However, we first focused our study on this toy network because it allows to easily identify the key biological features that are essential for the implementation of greediness.

The sketch of the simple motif is reported in Fig. 5a: two parallel pathways, here condensed in reactions A and B, convert precursor *P* in to metabolite *M* with stoichiometry  $1:\nu_A$  and  $1:\nu_B$ , respectively (i.e.  $P \xrightarrow{A} \nu_A M$  and  $P \xrightarrow{B} \nu_B M$ ). Without loss of generality we assumed reaction A is more efficient than reaction B (i.e.  $\nu_A > \nu_B$ ). Since *M* is used for the biomass production, this implies also that the growth rate recovery induced by resiliencing reaction A is smaller than the growth rate recovery induced by the resiliencing of reaction B ( $g_A < g_B$ , see Supplementary Fig. S9). The two negative feedbacks of the metabolite *M* on reactions A and B (e.g. allosteric effect) have been described through classical Hill equations with constants  $H_A$  and  $H_B$ , respectively. This led to the following set of equations:

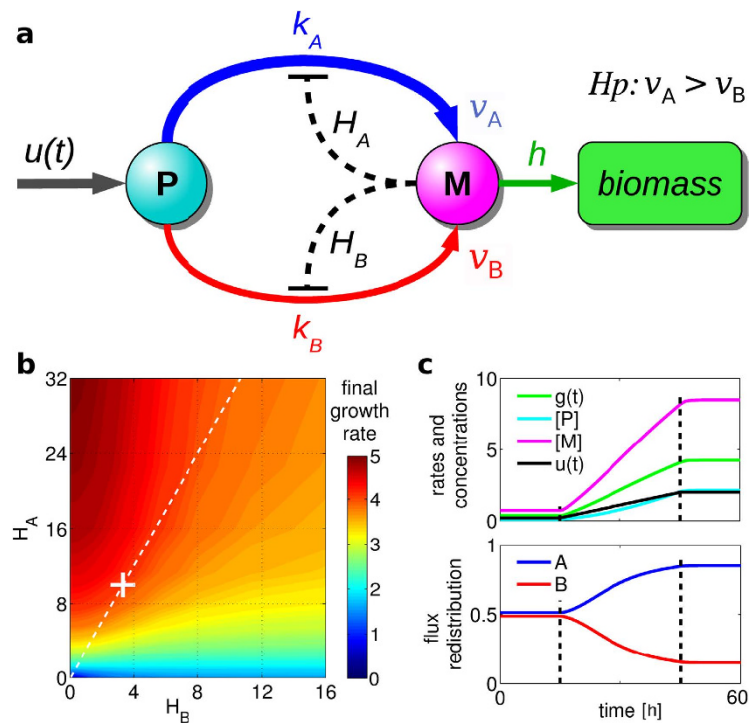
$$\begin{cases} \frac{d[P]}{dt} = u(t) - k_A \frac{H_A^2}{H_A^2 + [M]^2} [P] - k_B \frac{H_B^2}{H_B^2 + [M]^2} [P]; \\ \frac{d[M]}{dt} = \nu_A k_A \frac{H_A^2}{H_A^2 + [M]^2} [P] + \nu_B k_B \frac{H_B^2}{H_B^2 + [M]^2} [P] - h[M]. \end{cases} \quad (3)$$

The growth rate is then given by  $g(t) = h[M]$ . We considered the case when a previous step in the regulation dynamics has caused an increase of the availability  $u(t)$  of the precursor *P*. Given such an input, we analysed how the two feedbacks may affect the response of the systems to this change. Figure 5b reports the biomass production obtained by running the simulations with different values of  $H_A$  and  $H_B$ . This result shows that, in order to induce an higher increase of the growth rate, it is sufficient that evolution had shaped the inhibitory effects of *M* by increasing  $H_A$  and/or decreasing  $H_B$  such that  $H_A > H_B$ . We would like to point out that, as mentioned above, in such a simple network the induced short-term increase of the growth rate coincides with a global advantage: this is the reason why this system reaches high biomass production.

It is also interesting to estimate the probability of resiliencing reaction A through its effective kinetic parameter in equation (3), i.e.  $k_A^{\text{eff}} = k_A H_A^2 / (H_A^2 + [M]^2)$ . Since the resiliencing of reaction A gives  $g_A = h[M]$ , this leads to:

$$\mathbb{P}[A] = 1 - \frac{k_A^{\text{eff}}}{k_A} = 1 - \frac{H_A^2}{H_A^2 + [M]^2} = \frac{g_A^2}{(hH_A)^2 + g_A^2}, \quad (4)$$

which, by choosing  $H_A$  such that  $hH_A = g_B$ , is identical to the second-moment formula (2) that we already proved to describe well the metabolic regulation (see Fig. 2a from which we can also see that  $m \simeq \beta/100$ , i.e.  $m = 2$ ). Indeed, simulations of the network motif (3) with  $H_A = g_B/h$  and  $H_B = g_A/h$  reproduces the resiliencing of reaction B, coherently with our greedy criterion (see white cross in Fig. 5b and corresponding dynamics in Fig. 5c; more simulations in Supplementary Fig. S9). This indicates that evolution could have modulated these values and thus shaped the metabolic regulation in order to implement the greedy strategy.



**Figure 5.** An example of a possible greedy regulatory motif. (a) Sketch of the network: the metabolite  $M$  has an inhibitory effect on the two enzymes that convert a molecule of the precursor  $P$  into  $\nu_A = 3$  or  $\nu_B = 1$  molecules of  $M$  by reaction  $A$  or by reaction  $B$ , respectively. The two inhibitions are described through quadratic Hill equations with different constants  $H_A$  and  $H_B$ , see equation (3). (b) Effect on the final growth rate of the Hill constants  $H_A$  and  $H_B$  (all other parameter values in Fig. S10). White dotted line indicates the set of points such that  $H_A/H_B = g_B/g_A$  (where  $g_A$  and  $g_B$  are the growth rates after resiliencing reaction  $A$  and  $B$ , respectively – see Supplementary Fig. S9). White cross corresponds to the values used for the simulation in panels c, namely  $H_A = g_B/h = 10$  and  $H_B = g_A/h = 3.3$ . (c) Above: Simulated dynamics of the growth rate  $g(t) = h[M]$  (green) and of the concentration of the two compounds,  $P$  (cyan) and  $M$  (magenta), for a given precursor availability  $u(t)$  as input (black). Below: Redistribution of the fluxes between the two parallel reactions  $A$  and  $B$  (same simulation as above). More simulations in Supplementary Fig. S9. An example of network in which greediness leads to a suboptimal end-point is presented in Supplementary Fig. S10.

As mentioned at the beginning of this section, with such a minimal model it was not possible to differentiate local from global information, i.e. the short-term advantage in the growth rate recovery was also a long-term advantage. Therefore, in order to have a greedy and myopic dynamics that ends in a suboptimal point, a more complex network has been built. Supplementary section “An example of network with suboptimal end-point” presents an example of such a network and confirms that the proposed regulatory motif based on negative feedbacks still guarantees the implementation of the greedy strategy also in case of a partial recovery of the growth rate.

## Discussion

Cases of sub-optimal adjustment of the metabolism have suggested a myopic strategy as a guide for the microbial response to external perturbations: in our hypothesis this strategy is the myopic greedy resiliencing. The description we obtained for the considered wide range of different gene deletions and different carbon sources suggests that this principle is fundamental. Indeed, our criterion provides an explanation for important experimental features both at the macroscopic level and at the molecular level also for non-trivial metabolic responses such as the appearance of multiple phenotypes (Figs 1–4). It is in fact reasonable to think of the regulatory machine as having already been shaped by evolution in order to favour the recovery of the growth rate. However, it is unlikely that evolution has led to a system capable of an optimal recovery from all possible external perturbations. Similarly, in the case of bacterial chemotaxis the unavoidable conflicting requirements that come from different and incompatible chemoattractant profiles have shaped the signalling pathway towards a compromised chemotaxis response<sup>24</sup>. In this perspective, the proposed criterion of a greedy and myopic regulation appears to be the most reasonable and natural solution found by evolution.

The analogy with the energy minimization problem over a corrugated landscape is also straightforward. Considering the growth rate as the opposite of a Hamiltonian, a metabolic adjustment can be viewed as the dynamics of a particle that is moving toward the minimal energy (see Materials and Methods and Supplementary Figs S11 and S12 for more details). In this perspective, the regulatory machine might be seen as an algorithm that searches for the global minimum. In fact, it is well known that, depending on the initial point and on the complexity of the energy landscape, even the best available algorithm cannot guarantee to find the optimal solution.



Because of this, it is reasonable to expect that also the microorganism may end in a local minimum (as seen for *GLU-tpi* and *GLU-pgi*).

Moreover, we have presented an example of a simple regulatory motif which might be used by the cell to adopt the greedy strategy (Figs 5 and S10). While doing this, we showed how the description provided by GRAM is equivalent to the description derived from classical Hill equations (see equation (4)). Therefore, although the greedy hypothesis has been formulated through a very different and general approach, the resulting GRAM method is consistent with the standard mathematical formalism used for modelling cell regulation. All these three similarities with consolidated approaches might be considered as additional support to the plausibility of our interpretation as the basic and fundamental principle.

We would also like to underline that, the proposed method does not require any pre-existing information about the regulations. Since the entire description is based only on the stoichiometry of the reaction network, the proposed mechanism establishes a new functional link between regulation and stoichiometry of the metabolism. Therefore, as pointed out by the case of multiple phenotypes appearance we had considered in the validation step, our interpretation might also contribute for a better understanding of the genotype-phenotype correlation. For the same reason, our method introduces a new use of FBA which may enlarge the spectrum of applications of this computational tool inside Systems Biology.

## Materials and Method

The reconstructed stoichiometric network of the central metabolism of *E. coli*<sup>25</sup> has been extended to incorporate the methylglyoxal pathway. Supplementary Fig. S2 reports a sketch and the main features of this network. Moreover, after having verified that the carbon source is the only limiting factor for the biomass production, the upper-bound of its corresponding exchange reaction has been rescaled to the measured Supply Uptake Rate (SUR)<sup>9</sup>. Values are plotted in panel C of Supplementary Fig. S2. This constraint was necessary in order to reproduce the experimental growing conditions and to provide the quantitative and absolute scaling of the computed growth rate with respect to the experimental values. The same procedure had been adopted in ref. 9. However, in order to avoid any influence on the redox balance of the metabolism, we did not set any constraint on the Oxygen Uptake Rate (OUR).

GRAM represents an improvement of what has been reported in ref. 18: the current version provides a more useful and necessary biological interpretation of the phenomenon we want to describe and explain. For completeness, we reported the entire formalism in the Supplementary Information. The code for GRAM and the model of equation (3) have been implemented in MATLAB 2013R using ILOG-IBM CPLEX 12.1 (under academic licence) for solving all linear and quadratic optimization problems. All results from GRAM are the average over a set of 500 independent simulations using  $\beta = 200$  h.

## References

- Usui, Y. *et al.* Investigating the effects of perturbations to *pgi* and *eno* gene expression on central carbon metabolism in *Escherichia coli* using <sup>13</sup>C metabolic flux analysis. *Microb. Cell Fact.* **11**, 1–16 (2012).
- Li, M., Ho, P. Y., Yao, S. & Shimizu, K. Effect of *sucA* or *sucC* gene knockout on the metabolism in *Escherichia coli* based on gene expressions, enzyme activities, intracellular metabolite concentrations and metabolic fluxes by <sup>13</sup>C-labeling experiments. *Bioch. Eng. J.* **30**, 286–296 (2006).
- Flores, S., Gosset, G., Flores, N., De Graaf, A. & Bolvar, F. Analysis of carbon metabolism in *Escherichia coli* strains with an inactive phosphotransferase system by <sup>13</sup>C labeling and NMR spectroscopy. *Metab. Eng.* **4**, 124–137 (2002).
- Hua, Q., Joyce, A. R., Palsson, B. Ø. & Fong, S. S. Metabolic characterization of *Escherichia coli* strains adapted to growth on lactate. *Appl. Env. Microb.* **73**, 4639–4647 (2007).
- Mahadevan, R. *et al.* The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264 (2003).
- Segré, D., Vitkup, D. & Church, G. Analysis of optimality in natural and perturbed metabolic networks. *Proc. Nat. Acad. Sci. USA* **99**(23), 15112–15117 (2002).
- Shlomi, T., Berkman, O. & Ruppin, E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc. Nat. Acad. Sci. USA* **102**(21), 7695–7700 (2005).
- Cornelius, S., Lee, J. & Motter, A. Dispensability of *Escherichia coli*'s latent pathways. *Proc. Nat. Acad. Sci. USA* **108**, 3124 (2011).
- Fong, S. S. & Palsson, B. Ø. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nature Gen.* **36**, 1056–1058 (2004).
- Motter, A., Gulbahce, N., Almaas, E. & Barabási, A. Predicting synthetic rescues in metabolic networks. *Mol. Sys. Biol.* **4**, doi: 10.1038/msb.2008.1 (2008).
- Lewis, N. *et al.* Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Sys. Biol.* **6**, doi: 10.1038/msb.2010.47 (2010).
- Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M. & Sauer, U. Multidimensional optimality of microbial metabolism. *Science* **336**, 601–604 (2012).
- Beg, Q. K. *et al.* Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc. Nat. Acad. Sci. USA* **104**, 12663–12668 (2007).
- Mahadevan, R., Edwards, J. & Doyle, F. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Bioph. J.* **83**, 1331–1340 (2002).
- Meadows, A. L., Karnik, R., Lam, H., Forestell, S. & Snedecor, B. Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. *Metab. Eng.* **12**, 150–160 (2010).
- Kim, J. & Reed, J. L. RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. *Gen. Biol.* **13**, R78 (2012).
- Nishikawa, T., Gulbahce, N. & Motter, A. Spontaneous reaction silencing in metabolic optimization. *PLoS Comp. Biol.* **4**, e1000236 (2008).
- Altafini, C. & Facchetti, G. Metabolic adaptation processes that converge to optimal biomass flux distributions. *PLoS Comp. Biol.* **11**, e1004434 (2015).
- Fong, S., Nanchen, A., Plasson, B. & Sauer, U. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *J. Biol. Chem.* **281**, 8024–8033 (2006).
- Hua, Q., Yang, C., Baba, T., Mori, H. & Shimizu, K. Response of central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. *J. Bacter.* **185**, 7053–7067 (2003).

21. Kabir, M. & Shimizu, K. Gene expression patterns for metabolic pathways in *pgi* knockout *Escherichia coli* with and without *phb* genes based on RT-PCR. *J. Biotech.* **105**, 11–31 (2003).
22. Nor Aini, A., Shirai, Y., Hassan, M. & Shimizu, K. Investigation on the metabolic regulation of *pgi* gene knockout *Escherichia coli* by enzyme activities and intracellular metabolic concentrations. *Malays. J. Microb.* **2**, 24–31 (2006).
23. Charusanti, P. *et al.* Genetic basis of growth adaptation of *Escherichia coli* after deletion of *pgi*, a major metabolic gene. *PLoS Genetics* **6**, e1001186 (2010).
24. Celani, A. & Vergassola, M. Bacterial strategies for chemotaxis response. *Proc. Nat. Acad. Sci. USA* **107**, 1391–1396 (2010).
25. Orth, J., Fleming, R. & Palsson, B. Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide. *EcoSal* **4**, 1–7 (2006).

## Acknowledgements

I would like to thank Stephen S. Fong for providing the experimental data used in Figure 1. I thank also Claudio Altafini and Rea Antoniou Kourounioti for useful discussion and Areejit Samal for his suggestions on method improvement. This work was financially supported by SISSA.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The author declares no competing financial interests.

**How to cite this article:** Facchetti, G. A simple strategy guides the complex metabolic regulation in *Escherichia coli*. *Sci. Rep.* **6**, 27660; doi: 10.1038/srep27660 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>